

Defining Strong State Accountability Systems

How Can Better Standards Gain Greater Traction?

A FIRST LOOK

By Eileen Reed, Janie Scull,
Gerilyn Slicker, and Amber M. Winkler

April 2012

Contents

Introduction	1
Situating This Report	3
Defining Strong State Accountability Systems	5
Summary of Pilot Results	10
Closing	15
ESEA Waivers Alter State Accountability Systems	16
Appendix A: State Accountability Principles in Full	18
Appendix B: State Accountability Profiles	26
California	26
Colorado	37
Florida	46
Indiana	56
Massachusetts	65
Ohio	78
Texas	89

The Thomas B. Fordham Institute is the nation's leader in advancing educational excellence for every child through quality research, analysis, and commentary, as well as on-the-ground action and advocacy in Ohio. It is affiliated with the Thomas B. Fordham Foundation, and this publication is a joint project of the Foundation and the Institute. For further information, please visit our website at www.edexcellence.net or write to the Institute at 1016 16th St. NW, 8th Floor, Washington, D.C. 20036. The Institute is neither connected with nor sponsored by Fordham University.

Introduction

As of March 2012, forty-five states and the District of Columbia have adopted the Common Core State Standards (CCSS).¹ While states should be lauded for strengthening their expectations for students,² nobody should expect standards adoption *alone* to drive student growth. Indeed, a sensible “theory of action” should undergird the CCSS and articulate how the standards can better gain traction in states.

That (admittedly simple) theory goes something like this: The state adopts strong standards, administers outcome assessments that reflect the standards’ rigor and intent, and puts into place a robust accountability system that supports, rewards, and sanctions, as needed, schools, students, teachers, and other adults. Districts handle curriculum, instructional materials, and professional development; administer diagnostic measures; and provide other resources to help teachers implement the standards with fidelity in classrooms. But the *state*—it bears repeating—is uniquely positioned to send the message to educators, parents, and the broader public that a quality education is valued within its borders and that it will advance judicious policies to help ensure that schools, students, and adults are held to account for student success. After all, strong standards nested within a flimsy state accountability structure will surely wither.

Thus the time for states to reboot their accountability systems is now. They are poised to take advantage not only of the opportunity that CCSS affords but also of that presented by the federal Elementary and Secondary Education Act (ESEA) waivers. Three-quarters of the states have now applied for federal waivers; eleven were approved by the Obama administration in February 2012, and twenty-six states (plus the District of Columbia) are pending consideration as of early April

¹ Minnesota adopted the CCSS for English language arts only, and thus is not counted in this tally.

² For the majority of states, the CCSS marks a vast improvement over their previous mediocre-to-awful academic standards in English language arts and mathematics. See Sheila Byrd Carmichael, Gabrielle Martino, Kathleen Porter-Magee, and W. Stephen Wilson, *The State of State Standards—and the Common Core—in 2010* (Washington, D.C.: Thomas B. Fordham Institute, July 2010), <http://www.edexcellence.net/publications-issues/publications/the-state-of-state.html>.

(the third waiver deadline is September 2012).³ Moreover, the draft reauthorization bills that are making their way through Congress would likely give states even wider latitude to design their own approaches to accountability.

Worn down by a federal law that attempts to dictate uniform accountability for all jurisdictions, states have now come forward with their own ideas of how to fix a broken K-12 system. Some states will seize the opportunity to advance student performance more aggressively than ever before; others, almost inevitably, will shirk their commitments—perhaps because they think accountability is wrong-headed, mean-spirited, or unnerving—and take the easy road. The latter is understandable; after all, people and organizations don't much like accountability—at least not for themselves. Active resistance and foot-dragging are to be expected, much more so than in response to standards and assessments alone. All the more important, then, to help accountability agents to stay honest and their systems to become—or remain—forward-thinking, pragmatic, and relentless in their pursuit of better outcomes for students. That's the goal of this paper.

But what should strong state accountability systems—those that help ensure that the common standards get taught (and learned!) in classrooms—look like in 2012-13 and beyond?

Once upon a time, we talked about standards as one leg of a three-legged stool: They must be accompanied by strong assessments that accurately measure student performance *and* by strong accountability systems that tie school- and district-level consequences to that performance. While those building-block elements still have merit today, they also need a fresh remodeling. Version 2.0 of state education-reform systems must broaden accountability requirements to include students and individual adults—not just the buildings and districts that they occupy. These new systems will also need to be nimble enough to embrace advances in how we measure student progress, teacher and principal effectiveness, and education productivity.

Though we don't yet know what this will look like, the CCSS will soon stand on its second leg: Two federally funded consortia of states are currently crafting new systems of annual assessments aligned to the common standards. (We hope that these tests will reflect the rigor and intent of the standards to which they are to be aligned.) Even with the standards and assessments legs of the stool in place,

³ In September 2011, the U.S. Department of Education established a waiver application process by which states can bypass some of the more onerous and objectionable aspects of NCLB. In exchange, states had to agree to implement a number of reforms relative to college- and career-ready standards, aligned assessments, differentiated accountability systems, and teacher and principal evaluation systems, among other areas. See *ESEA Waivers Alter State Accountability Systems* on page 16 for more.

however, the effectiveness of the CCSS in each state will depend greatly on the strength of each state's accountability system.

Leaning on our own research and that of others, this report lays out the major principles of a strong state accountability system and our rationale for their inclusion. We follow that with a short discussion of common accountability pitfalls and an in-depth analysis (in Appendix B) of seven state systems that illustrate the strengths and challenges of developing these systems. Our key question: What accountability policies can states put in place to support the efficacy of the new common standards and assessments, and what should they be looking to overhaul or improve?

Situating this Report

This report is a pilot study intended to inform a larger analysis of the accountability systems in every state (and the District of Columbia) during the early years of Common Core implementation. We ask that the reader treat it as such and provide us with feedback on the accountability principles contained herein.⁴ We plan to apply these principles, once revised, to all fifty state accountability systems in order to appraise their quality. Our first national report is slated for early 2013, with follow-up studies two and four years later. Tracking systems in this manner will prove beneficial because many states will be in “flux” over the next several years as they refine and adapt their systems based on the demands of the Common Core and on the plans and promises outlined in their recently approved waivers (and/or those provisions detailed by ESEA reauthorization legislation—assuming Congress one day gets its act together).

Fordham is also conducting three other studies pertinent to CCSS implementation. The first is an analysis of Common Core implementation costs; the second, an in-depth study of district-level implementation of CCSS; and the third, a nationally representative survey of English language arts teachers that assesses the rigor of their reading assignments both before and after implementation of CCSS (summer 2012 and spring 2015).⁵

Acknowledgments

Many thanks go to Eileen Reed, educational consultant, and Gerilyn Slicker, George Washington University graduate student and 2012 TFA corps member, who

⁴ Feedback can be emailed to accountability-metric@edexcellence.net.

⁵ See also previous Fordham Institute publications relevant to the Common Core, including *The Common Core and the Future of Student Assessment in Ohio* (August 2011); *Now What? Imperatives and Options for Common Core Implementation and Governance* (October 2010); *The State of State Standards—and the Common Core—in 2010* (July 2010); *Common Education Standards: Tackling the Long-Term Questions* (June 2010); *Review of the Draft K-12 Common Core Standards* (March 2010); and *Stars by Which to Navigate? Scanning National and International Standards in 2009* (October 2009).

conducted much of the background research and drafted an early version of the report. At the Fordham Institute, Amber Winkler, vice president for research, and Janie Scull, research analyst and production manager, updated the content and edited the final draft.

Many individuals also provided feedback or assistance during the course of this study. For our seven state accountability profiles, we interviewed state and district representatives in each state, as well as stakeholders working outside of the education system. We also conducted interviews with representatives from a handful of national-level education organizations. Because these interviews elicited candid commentary on both the structure and functionality of the state accountability systems, we keep these names anonymous. In addition, we solicited feedback from a number of national and local stakeholders on our draft metric, including, but not limited to: Kerri Briggs (George W. Bush Institute), Michael Cohen (Achieve), Linda Noonan (Massachusetts Business Alliance for Education), Alissa Peltzman (Achieve), Elliot Regenstone (Education Counsel), and Judith Rizzo (Hunt Institute).

Generous support for this project was provided by the Bill & Melinda Gates Foundation, as well as by our sister organization, the Thomas B. Fordham Foundation. A big thank you goes out to the Fordham team for assistance on this project, especially Chester E. Finn, Jr., Michael J. Petrilli, and Kathleen Porter-Magee, for their project guidance and feedback, and to Tyson Eberhardt and Joe Portnoy for dissemination.

Defining Strong State Accountability Systems

We're not the first to delineate a framework for sound state accountability systems. Our thinking was informed by the good work of several other organizations, including the Council of Chief State School Officers (CCSSO) and the George W. Bush Institute. CCSSO's *Roadmap for Next-Generation State Accountability Systems* lays out nine principles—some of which we adopt verbatim—relative to standards, student outcomes, annual determination of school and district performance, timely data, and interventions for failing schools, among other areas.⁶ Similarly, in *Advancing Accountability*, the Bush Institute delineates ten school-level accountability principles for states, including provisions pertaining to performance and content standards, assessments, reporting, intervention, and choice alternatives.⁷ The primary difference between those excellent reports and this one is the focus on *who* is to be held accountable. Prior reports primarily target schools and districts, while we broaden the focus to include individuals: students, teachers, principals, and other adults. After all, smart accountability should start with the individual. Not that collective accountability has no value; rather, it should begin with responsible individuals who also “have skin in the game.” (See sidebar, *Putting Our Cards on the Table*, for more of our assumptions.)

Thus, we posit that strong state accountability systems have six essential elements:

1. Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards;
2. Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement;

⁶ See Council of Chief State School Officers, *Roadmap for Next-Generation State Accountability Systems* (Washington, D.C.: Council of Chief State School Officers, 2011), http://www.ccsso.org/Resources/Publications/Roadmap_for_Next-Generation_State_Accountability_Principles.html.

⁷ See The George W. Bush Institute, *Advancing Accountability* (Dallas, TX: The George W. Bush Institute, February 2012), http://www.bushcenter.com/blog/wp-content/uploads/2012/02/Advancing_AccountabilityFeb-2012.pdf.

3. Annual determinations and designations for each school and district that meaningfully differentiate their performance;
4. A system of rewards and consequences to drive improvement at the school and district levels;
5. A system of rewards and consequences to drive improvement at the individual student level; and
6. A system of rewards and consequences to drive improvement at the individual teacher and administrator level.

Below, we briefly describe each; for detailed explanation and scoring guidelines, see Appendix A.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

Sound content standards are an essential *starting point* for improved student outcomes. States should, at a minimum, have them in place for English language arts (including writing), mathematics, U.S. History, and science, and should assess student mastery of subjects at regular intervals. Further, states should hold all students to high standards by setting rigorous cut scores for each test.⁸

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Robust and timely diagnostic and outcome data on school, student, and adult performance enable educators, policymakers, parents, students, and the broader public to take appropriate action on the decisions that concern them (e.g., teacher hiring, school choice, classroom interventions, selection of turnaround model, just to name a few). Though we appreciate the need for a wide variety of data to assess sundry outcomes at all levels, we target here the data of most concern: student performance.

States should release annual aggregate student outcome data in an accessible manner (meaning clear and user-friendly). In short, states should publish results for every school that includes a snapshot of how many students are reaching various achievement levels (such as basic, proficient, advanced, etc.), how many students are making expected progress over time, and whether or not the school

⁸ Note that proper alignment of standards and tests is critical—including a matchup of rigor—but we're unable to evaluate that alignment.

Putting Our Cards on the Table

The metric that we outline here is premised on certain assumptions and faces some limitations. First, the principles we discuss are meant to help states define the *parameters* for strong accountability systems. We don't pretend that there is one best way to design these systems. Even if there *were* one best approach today, the constantly changing landscape requires that states persist in experimenting with and tweaking their systems. States continue to scale the steep learning curve erected since the establishment of No Child Left Behind: Timid sanctions and unimplemented interventions leave many questioning how best to address failing schools; local stakeholders remain befuddled in their attempts to navigate dual federal and state accountability systems (though the waivers should help on this front); unexplained student growth or stagnation begs additional examination; and so on. Thus, many states are experimenting with innovative methods to hold schools, adults, and students accountable for outcomes; to gather and report data more usefully; to differentiate among dismal, failing, average, good, and great schools; and to incentivize and crack down on particular behaviors. Our principles are therefore more holistic than prescriptive, leaving ample room for experimentation and innovation. For example, we are agnostic as to how states choose to incentivize effective leadership—but they should recognize the importance of such policies.

Second, measuring the *quality* of state policies is easier said than done, and we do not attempt to do so here. We are well aware, for instance, that simply deploying a high school graduation test for the purpose of student-level accountability does not mean that the test is any good. It may tell us nothing about whether or not a passing grade indicates a student is ready for college or career. But to answer the “quality question” would require more time and hands than our resources allow, as metrics by which to evaluate the usefulness, rigor, and application of particular policies simply do not exist. For instance, we favor states that require individual students to pass “gateway” assessments, perhaps in elementary reading, before being promoted to the next grade. But to evaluate the application of that policy we'd also need to know how many students failed the gateway exam in a given year and were subsequently retained as a result. Few states release these data in timely fashion, if at all. So states can have a policy on the books but not enforce it—or not measure its enforcement—and we do not account for such lax implementation in our metric.

Third, we concentrate on provisions relative to accountability and incentives rather than those intended to support failing schools and students or weak teachers and principals. Our focus is on ensuring robust data, making results transparent, rewarding good work, and penalizing chronically low performance. We recognize that providing tailored support to organizations and individuals in need is critical to a strong accountability system, but assessing the adequacy of these measures is difficult: Many states offer similar supports of the same name and it is nearly impossible to draw meaningful distinctions among them without in-depth analysis.

Finally, a clarification: While we recognize that the term “intervention” can take one of two forms—supports (e.g., state-operated regional service centers or technical assistance in improvement planning) or consequences (e.g., mandatory replacement of staff or contracting with outside entities to run the school)—we address in these pages only the latter.

itself is improving from year to year. (Of course, states should also require that schools provide reports to parents on their own child's performance and progress.) In addition, the state should collect and report data on multiple measures of student performance whenever possible. These might include attendance, graduation rates, on-track indicators, postsecondary matriculation/retention or remediation rates, course completion (transcript data), attainment of industry certification, and performance on SAT, ACT, and AP/IB exams. State databases should also permit school-level comparisons.

States should also require that the data above be disaggregated by subgroups within schools of race/ethnicity, gender, free and reduced-price lunch status, special-needs status, English language learner status, and so on. States might also choose to substitute one or more of these groups with a low-performing subgroup or high-performing subgroup (i.e., a "super subgroup").

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

Stakeholders need to know how schools and districts compare with one another. States need to characterize and differentiate among schools and districts based on student achievement in a valid and reliable manner.

These designations should not obscure the truth (i.e., via inflated performance or categories that permit high percentages of schools to fall disproportionately into the top tiers). They should be user-friendly—we prefer the A-to-F grading system. And the state should integrate state and federal accountability systems so that designations do not conflict.⁹

School and district designations should be derived, at least in part, from measures of individual student growth (either normative growth or growth to standard). Those designations should also be informed by the performance of specific student groups (e.g., race/ethnicity, income, English language learner status, and special-needs status) or "super subgroups."¹⁰

4) A system of rewards and consequences to drive improvement at the school and district levels

Strong state accountability systems engage the multiple entities involved in educating children, including districts, schools, educators, and students, and incorporate both carrots and sticks to spur positive action and behavior. At the

⁹ As mentioned, the federal ESEA waivers are encouraging this integration.

¹⁰ States might also want to use some form of human judgment when rating schools. The "school inspectorate" idea, borrowed from the United Kingdom, is particularly intriguing. Such inspectors could potentially adjust a school's letter grade up or down depending on what they observe on the ground.

district and school level, incentives may include “honor roll” or similar “blue ribbon/distinguished” status, financial rewards, increased autonomy in operations/spending, and other regulatory relief (e.g., automatic renewal of district accreditation) in exchange for meeting heightened district accountability provisions and/or as a reward for performance or growth.

State accountability systems should also require targeted interventions into low-performing schools.¹¹ These may include replacement of staff, charter conversion, state takeover, contracts with outside entities to operate schools, and automatic school closure after consecutive years of failure. (Interventions may also be catalyzed via “parent trigger” provisions.) Such consequences need to be *real* (or have “teeth”), so the state’s system should avoid introducing loopholes by which low-performing schools or districts can evade tough sanctions; water down or delay sanctions; and/or restrict sanctions to a subset of the lowest-performing schools.¹²

5) A system of rewards and consequences to drive improvement at the individual student level

In the last decade, we’ve spent much time thinking about how to hold districts, schools, and—less often—educators accountable. But students are often not held to account for their own learning. Strong state accountability systems put some of the responsibility for successful student outcomes back where it belongs—on the shoulders of individual students, and, implicitly, their parents, guardians, and others who influence their actions and attitudes.

Consequently, states should have on the books multiple methods by which individual students are held to account for their performance. These may include, but are not limited to: requirements that students pass cumulative high-school exit exams in the core subject areas in order to receive diplomas; pass end-of-course exams to get credit for courses or pass end-of-course exams that are integrated into final course grades; and pass “gateway” assessments in one or more key grades and subjects in order to be promoted to the next grade (e.g., pass a reading test at the end of third grade). States may also have a “no pass, no play” policy that requires students to pass all of their courses in order to participate in extracurricular activities, or a “no pass, no driver’s license” policy. (We’re open to other ideas here, too.)

¹¹ As discussed on page 7, interventions comprise both supports and sanctions. We’ve targeted the latter here.

¹² As mentioned, we’re unable to assess quality in this metric as thoroughly as we’d like. Still, we’ve attempted here to illustrate what we think good state policy does and does not do when it comes to meaningful interventions for low-performing schools.

Finally, states might also reward individual students via guaranteed entry into a state college and/or guaranteed college financial aid (or “priority status”) for eligible students with high GPAs or high SAT or ACT scores.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

States should require annual evaluations of all teachers. Further, teacher evaluations for all core subject areas should include measures of student performance and/or growth on state assessments—and those results should inform decisions related to continuous employment.

The state should also have in place policies that incentivize individual teacher effectiveness and productivity, such as merit pay and/or other individual performance-based bonuses (including for teaching in high-needs schools), extended contracts, career ladders (e.g., master teachers), personnel decisions based on performance (e.g., no “last in, first out” policies), and so on.

Adult accountability does not end with teachers. The state should also mandate annual evaluations of all principals and these should be based, in part, on their schools’ performance; results should inform employment status. Finally, the state should offer incentives and rewards for effective principals, including bonuses, extended contracts, enhanced autonomy, and so on. States might also offer incentives for superintendents and/or school board members based on their schools’ performance.



If lawmakers can put these six elements in place, they’ll be well on their way to bolstering the likelihood of success for students, educators and other adults toiling away in the state’s schools. Let’s turn now to the findings in some of these states.

Summary of Pilot Results

As indicated, our work builds upon the prior research of other organizations. But it is also largely informed by in-depth analysis of the accountability systems in seven states: California, Colorado, Florida, Indiana, Massachusetts, Ohio, and Texas. (See sidebar, *Selection of States*, for why we chose this set.)

Each of the seven detailed profiles can be found in Appendix B. We conducted these analyses before the first batch of states applied for ESEA waivers, so some of the information for four of our states (Colorado, Florida, Indiana, and

Selection of States

We chose states for this pilot study based in part on laudable properties of their state accountability systems, but more broadly on their strong history of education reform. We reasoned that aggressive reform states might also be prime candidates for strong accountability systems. Here's a sampling of the elements that informed our selection.

California and **Massachusetts** have long been recognized for their strong academic standards,¹³ though the Golden State cannot boast the impressive performance of the Bay State, which routinely scores at the top of the heap domestically and near the top internationally.¹⁴ Massachusetts is well-known and respected for setting rigorous passing standards on state assessments and for requiring that students meet those standards to earn a diploma.

Florida and **Indiana** have been recognized for their high-quality state assessments.¹⁵ The Sunshine State has also posted impressive gains for student subgroups, particularly its Hispanic students.¹⁶ A perennial leader in school innovation since Governor Jeb Bush launched reforms in 1999, Florida passed a package of reforms in spring 2011 that abolished tenure for newly hired teachers, established a performance-based pay system, provided additional pay for high-need subject areas and at-risk schools, and further expanded charter schools and digital learning. Likewise, the Hoosier State reformed collective bargaining in 2011, expanded charter campuses, removed caps on virtual schools, and passed an expansive voucher program for low-income students.

In 2003, **Ohio** was one of the first states to mandate that value-added models be used to measure student performance. The Buckeye State also boasts two voucher programs and a robust school choice climate. In 2006, the state instituted an automatic closure ("death penalty") provision for poorly performing charter schools, considered to be the toughest closure law in the nation (a similar policy does not exist for traditional schools).

Texas was one of the first states to shine a light on the performance of all student groups—a concept that then-Governor George W. Bush sought to enshrine in No Child Left Behind when he became president. The Lone Star State has posted impressive gains for minority students

¹³ See the Fordham Institute's recent reports on state standards, *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, and *The State of State Science Standards 2012*, at <http://standards.educationgadfly.net/>; "Quality Counts 2011: Uncertain Forecast—State Report Cards," *Education Week*, January 2011, <http://www.edweek.org/ew/qc/2011/16src.h30.html>; and American Federation of Teachers, *Sizing Up State Standards 2008* (Washington, D.C.: American Federation of Teachers, 2008), <http://www.aft.org/pdfs/teachers/sizingupstandards0308.pdf>.

¹⁴ For example, while Massachusetts eighth graders scored higher than any other state's eighth graders on the 2009 reading and mathematics National Assessment of Educational Progress (NAEP), California's eighth graders scored third-to-last among the states in reading and fourth-to-last in mathematics. See the NAEP website at <http://nces.ed.gov/nationsreportcard/>. See also Gary Phillips, *International Benchmarking: State Education Performance Standards* (Washington, D.C.: American Institutes for Research, October 2010), http://www.air.org/files/AIR_Int_Benchmarking_State_Ed__Perf_Standards.pdf.

¹⁵ See the Florida and Indiana profiles in "Quality Counts 2011: Uncertain Forecast—State Report Cards," *Education Week*, January 2011, <http://www.edweek.org/ew/qc/2011/16src.h30.html>.

¹⁶ Matthew Ladner and Dan Lips, "Demography and Destiny? Hispanic Student Success in Florida," *Education Next* 9, no. 3 (Summer 2009):20-27, <http://educationnext.org/demography-as-destiny-2/>. See also the NAEP website at <http://nces.ed.gov/nationsreportcard/>.

(though these have flattened of late) as well as set the standard for the country to test students regularly in the core content areas, disaggregate performance, and tie accountability measures to those outcomes.

Colorado is recognized for its robust system of school choice. Its capital city also claims a sophisticated teacher evaluation and compensation system, one of the nation's first pay-for-performance pioneers. In addition, in May 2011 it passed a comprehensive legislative package (S.B. 191), which reformed teacher tenure, defined teacher effectiveness based primarily on student performance, and permitted districts to lay off teachers based on performance rather than seniority.

Massachusetts)¹⁷ may now be outdated. (See the sidebar *ESEA Waivers Alter State Accountability Systems* for a discussion of overall patterns in the approved waivers). We gathered information on state accountability systems from multiple sources, including state department of education websites; interviews with present and past employees of the state education agencies,¹⁸ district superintendents and/or other district staff, and on-the-ground stakeholders; and extant data sources.¹⁹ Interviewing multiple insiders at the state and local levels provided a window into how state- and district-level accountability policies and practices interacted (and sometimes collided) on the ground. In addition, we reached out to representatives of national organizations—including the Council of Chief State School Officers, the U.S. Education Delivery Institute, Achieve, and the Aspen Institute—in order to gain a better understanding of the current accountability landscape and add context to our state reviews. In total, we interviewed over forty education representatives between summer 2011 and early winter 2012.

In short, our examination yielded six themes—three key strengths in state accountability systems and three key weaknesses. First, the strengths:

1. **A number of states have developed clear and comprehensive means for rating district and school performance.** Robust accountability systems turn multiple data points into user-friendly labels. For instance,

¹⁷ Ohio has applied for a waiver in the second round of requests.

¹⁸ Texas is the one exception—the Texas Education Agency declined our request to be interviewed for this paper, and thus the Texas profile includes district-level, but not state-level, input. For Massachusetts, we interviewed personnel associated with both the Executive Office of Education and the Department of Elementary and Secondary Education.

¹⁹ In particular, we consistently drew information from the National Council on Teacher Quality's 2011 *State Teacher Policy Yearbook*. Other sources are cited in the individual profiles.

Florida and (more recently) Indiana use A-to-F grades to evaluate performance. These identifiers are informed by an array of achievement measures, including student proficiency and growth—but also by such other indicators as graduation and high school dropout rates, attendance, participation in AP/IB courses, achievement gaps, and performance on the SAT and ACT. Florida, for example, includes in its measures the performance of the lowest 25th percentile of students at each school, while Texas considers schools' college-going rates and college-level remediation rates.

2. **States are collecting a variety of data and making progress toward rendering them more “transparent” and “user-friendly.”** States must provide information to educators, parents, and taxpayers in order to help them understand and evaluate school and district results. And indeed, many states ensure that such stakeholders are awash in data relative to state, district, and school performance as well as sundry other areas (see above). Some states also publish results from various diagnostic and progress monitoring tools, as well as discrete data from online courses. Presenting these myriad data in a user-friendly fashion is a continuing challenge, but states such as Ohio are making inroads. In addition to its school report card, Ohio also issues a web-based interactive Local Report Card (iLRC) that easily allows comparison among schools and districts.
3. **Teacher accountability is improving.** More and more states, such as Florida, Indiana, and Colorado, are requiring student achievement measures to feature prominently in teacher evaluations, which then inform employment decisions. More states are also questioning automatic tenure after two to three years and experimenting with various performance-pay options.

Now, for the weaknesses:

4. **States still struggle with how to meld state and federal accountability systems.** Since the introduction of the No Child Left Behind Act, states have grappled with how best to integrate state and federal requirements so as to avoid conflicting messages. Of course, the Obama administration's waivers are intended to help address this problem. We hope that, in time, ESEA reauthorization will address it as well. While a few states have seamlessly combined their federal and state systems, most struggle to operate two parallel, disjointed systems—resulting in dual performance standards, designations, and sanctions for schools and districts. Such a twofold structure is not just onerous for schools and districts, but also confusing for educators and parents, particularly when a school meets one system's performance standards but not the other.

Massachusetts is an exemplar here. It designed one system that integrates federal and state accountability provisions. This allows the Bay State to report one marker of school performance based on *both* state and federal measures, and to set individual growth goals that also meet federal requirements for each school and district.

5. **Most states have only weak incentives to drive school, district, principal, and student performance.** Generally, states have policies that require intervention on the behalf of low performers—whether they are districts, schools, or students—but few policies that incentivize or reward high performers. Further, recent budget reductions have largely eliminated performance-based incentives for schools, teachers, and principals. (Accountability policies, in fact, tend to ignore positive incentives for school leaders altogether.) States do a *tad* better rewarding students; a couple of them offer high-performing students automatic college admission or financial aid.
6. **Sanctions are non-existent or ineffective.**²⁰ We're big supporters of transparent data both to motivate and to shame schools, districts, and individuals to action. But transparent data must be coupled with tangible incentives that reward high performance (as in #5 above) and meaningful sanctions that penalize low performance. Previous research has shown (and those we interviewed agreed) that NCLB's cascade of interventions was unsuccessful, in part, because districts took the road of least resistance. How might low performance be handled differently this time around? There's heated discussion about the capacity of state education departments to turn around low-performing schools—whether that be 5, 15 or 50 percent of them—but there's also a need to think more fundamentally about which approaches are right for which schools and what entities should be charged with overseeing change. State education departments, districts, business communities, parents, nonprofits, for-profits, and other entities likely all have a role in helping struggling schools, students, and adults. But very few states are recruiting and equipping multiple partners in the challenge.

²⁰ Again, “sanctions” implies the use of the stick, not the carrot. We do not include here state efforts to offer supports to schools.

Closing

States have two historic opportunities before them that could improve the odds for kids. First, the advent of the Common Core standards gives them the chance to repair two legs of their tripod: the academic standards and the assessments aligned to them. Most states have already traded up their subpar state standards in exchange for the CCSS. Our fingers remain crossed on the assessments.

The administration's waiver process offers the second opportunity. It allows states to restore the third leg of their tripod—the accountability limb. They can take advantage of this opening to infuse their schools with demanding content standards and rigorous assessments; collect, use, and report actionable data to all stakeholders; develop user-friendly classification systems that differentiate school and district performance; and devise novel ways to reward and reprove schools, districts, students, and adults. Or they can squander their chance by rolling back or watering down the tougher accountability provisions, masking the performance of low performers, and/or overlooking the commendable progress of hardworking students and adults.

We stand at a critical juncture. Will states make the most of the opportunities before them? Or will they recoil and renege on their responsibility to students, parents, and communities? We'll follow this unfolding story over the next several years, and let you decide.

ESEA Waivers Alter State Accountability Systems

Just before this report went to press, the Department of Education granted federal waivers to eleven states, allowing them to circumvent some of the requirements in the Elementary and Secondary Education Act (ESEA). Four of the eleven states are profiled in this report: Colorado, Florida, Indiana, and Massachusetts. (Ohio has applied for a waiver in the second round of requests, while California and Texas have not.) The new regulations outlined in the waivers dramatically alter existing accountability structures in some states and only tweak them in others. They do not, however, impact the essential accountability elements that we outline in this report.

To be eligible to apply for waivers, states had to address three broad principles outlined by the Department of Education. States must: (1) establish college- and career-ready standards and assessments; (2) develop accountability systems that differentiate school performance, rewards, and interventions; and (3) establish teacher/principal evaluation systems based in significant part on student achievement. The findings in this report are most affected by the second and third provisions, though surely the first one has an effect on state accountability systems as well, as we discuss in the Introduction to this paper. Though we did not have adequate time to revise our state profiles to reflect the new provisions outlined in the states' accepted waiver applications, the Center on Education Policy (CEP) *has* examined them; drawing from their research, we distill a number of general trends below.²¹

New provisions better integrate federal and state accountability. The most predictable trend is that the eleven states with ESEA waivers have designed systems to better streamline federal and state accountability regulations. As noted in a number of profiles in this report, dual state and federal systems have proven redundant, confusing, and burdensome. The waivers allow for more flexibility in setting annual objectives and intervening in low-performing schools.²² Some of the eleven states have completely revamped their accountability systems as a result; others, such as those currently operating under *Differentiated Accountability* systems (Florida and Indiana, in this report), have outlined new provisions that are more closely aligned with their current systems.²³ It is also important to note that, under the waivers, most of the eleven approved states have committed to apply accountability provisions to *all* schools, not just Title I schools.

Annual Measurable Objectives (AMOs) are no longer built on the goal of universal proficiency by 2013-14 and are no longer tied to accountability actions. For all eleven states, AMOs will no longer be based on AYP and the goal of reaching 100 percent proficiency by 2013-14. Instead, Florida has committed to reduce its number of non-proficient students by 50 percent within six years (and Massachusetts has opted for a more complex variation of this goal). Indiana has determined that all schools will receive a grade of A, or improve by two letter grades, by 2019-20. And Colorado aims to move all students to achieve at 2010's 90th percentile by 2015-16.

In many cases, these AMOs will differ by group, grade, and/or school. Some states have further expanded AMOs to include measures beyond static performance: Colorado and Florida will include additional data on

²¹ See "Major Accountability Themes of Approved State Applications for NCLB Waivers," Center on Education Policy, February 27, 2012, http://www.cep-dc.org/cfcontent_file.cfm?Attachment=Riddle%5FPaper%5FWaiverApp%5Fo22712%2Epdf.

²² CEP reports that "school districts receive little attention in the accountability provisions in the 11 state waiver plans...", though some states—including Colorado and Massachusetts—still plan to include annual objectives for states and report their progress against them. *Major Accountability Themes*, 14.

²³ The Department of Education approved *Differentiated Accountability* models for nine states in 2008 and 2009. In short, *Differentiated Accountability* allows states some flexibility in aligning state and federal accountability systems to reduce redundancy and confusion. The state and federal structures continue to operate independently of each other, but states may streamline what schools and districts are asked to do under both systems so that they are implementing cohesive strategies—not overlapping or conflicting ones.

measures such as student growth, achievement gaps, and college- and career-readiness. But now, many states will *not* use AMOs to determine accountability actions, as AYP was used. Instead, AMOs will primarily be used for reporting and/or diagnostic purposes, while other measures will be used to differentiate school performance and determine accountability actions.

Performance categories, not AMOs, will determine accountability actions and will include fewer subgroups than AMOs. Most states will establish multiple performance categories for schools and districts based on measures more complex than AYP. These measures, rather than AMOs, will be used to direct accountability actions. This should come as no surprise, as nearly all the states profiled herein already layer more detailed state-determined performance categories and consequences on top of school and district AYP status. (The four states in our report that have received waivers already have such performance categories in place.) As described in at least eight of the eleven approved waiver applications, these performance categories will be based on a combination of student performance, growth, and college-readiness data, such as dropout and graduation rates and results of SAT, ACT, and AP tests. Performance categories will be used to determine accountability actions.

Further, at least seven of the eleven states will base performance categories on only two subgroups—all students, and a single “disadvantaged subgroup”—rather than on the multiple subgroups used to determine accountability actions under AYP. In Florida and Indiana, for example, the “super subgroup” will consist of the lowest-performing 25 percent of students. In Massachusetts, it will combine low-income students, students with disabilities, and English language learners. The performance categories will be used to identify high achievers for rewards and low achievers for interventions. Only once low performers have been identified will data pertaining to subgroups be used to inform the type of interventions that might be undertaken.

To be eligible to receive waivers, states had to commit to identifying both Priority schools—the lowest-performing 5 percent (or more) of schools—and Focus schools—the 10 percent (or more) of schools with the largest gaps in achievement and graduation rates among specific subgroups.²⁴ Priority schools require the most intensive interventions, while Focus schools require some targeted interventions. To meet these regulations, Florida will designate schools with F grades as Priority schools, and those D grades as Focus schools. Indiana schools receiving Fs, or Ds and Fs for two consecutive years, will be Priority schools, while all other schools receiving Ds will be Focus schools. And in Massachusetts, Level 4 and 5 schools will be deemed Priority schools, while Level 3 schools will be Focus schools. In addition to Priority and Focus schools, states must identify high-performing Reward schools for recognition (and many states have committed to provide additional funding to these schools, contingent on future budgets). Most states will designate two types of high-performing Reward schools: one based on absolute performance and one on growth.

In short, while ESEA waivers have changed the game for states struggling under burdensome federal regulations, the waivers do not alter the general structure of a sound accountability system. Rather, the essential elements—such as performance differentiation and a system of rewards, consequences, and support—remain just as essential. While four of the systems used to inform our metric (and described in our profiles) are now different, they still serve to illustrate many of the strengths—and weaknesses—of state accountability systems.²⁵

²⁴ Waiver regulations only require that Priority schools total the lowest-performing 5 percent *of Title I schools*, and Focus schools total the 10 percent *of Title I schools* with the largest gaps in achievement and graduation rates. But, as mentioned earlier, most of the eleven waiver states have committed to applying their accountability systems to all schools, not just their Title I schools.

²⁵ As we begin to evaluate all fifty states against our accountability metric, we will re-review the states profiled here that have since received NCLB waivers.

Appendix A: State Accountability Principles in Full, with Scoring Ranges

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

- A. State content standards in English language arts (ELA), mathematics, U.S. History, and science are rigorous, clear, and specific.²⁶

Strong: Standards earn an A or B average across the subjects.

Medium: Standards earn a C average across the subjects.

Weak: Standards earn a D or F average across the subjects.

- B. State assesses student learning in the core content areas (ELA with writing, mathematics, history, and science) to determine student mastery of state content standards

Strong: State fulfills the federal requirements for ELA and math (testing annually in third through eighth grades, plus once in high school), and assesses science annually and U.S. History at least once in each of three grade spans (elementary, middle, and high school).

Medium: State fulfills the federal requirements, plus either tests annually in science or tests at least once in each of three grade spans in U.S. History.

Weak: State fulfills only the federal requirements (ELA and math assessments annually in third through eighth grades, plus once in high school; science assessments once in each of three grade spans).

²⁶ As measured by the Fordham Institute's recent reports on state standards: *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, and *The State of State Science Standards 2012*. State results can be found at <http://standards.educationgadfly.net/>.

C. State has in place rigorous assessment cut scores.²⁷

Strong: The state's proficiency cut scores equate to the **proficient** achievement level of the National Assessment of Educational Progress (NAEP) for all of the following:

Grade 4 Reading

Grade 8 Reading

Grade 4 Mathematics

Grade 8 Mathematics

Medium: The state's proficiency cut scores at least equate to the **basic** achievement level of the NAEP for all of the following:

Grade 4 Reading

Grade 8 Reading

Grade 4 Mathematics

Grade 8 Mathematics

Weak: The state's proficiency cut scores equate to the **below basic** achievement level of the NAEP in one or more of the following:

Grade 4 Reading

Grade 8 Reading

Grade 4 Mathematics

Grade 8 Mathematics

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

A. State releases annual aggregate student outcome data in an accessible manner (meaning clear and user-friendly) and state databases permit school-level comparisons.

Strong: For every school, states should publish results that include a snapshot of how many students are reaching various achievement levels (such as basic, proficient, advanced, etc.), how many students are making expected progress over time, and whether the school itself is improving from year to year. (Of course, states should also require that schools provide reports to parents on their own child's performance and progress.) In addition, the state should collect and report data on multiple measures of student performance whenever possible. These might include attendance, graduation rates, on-

²⁷ As measured by the National Center for Education Statistics' report, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>.

track indicators, postsecondary matriculation/retention or remediation rates, course completion (transcript data), attainment of industry certification, and performance on SAT, ACT, and AP/IB exams. State databases should also permit school-level comparisons.

States should also require that the data above be disaggregated by school-level subgroups of race/ethnicity, gender, free and reduced-price lunch status, special-needs status, English language learner status, and so on. States might also choose to substitute one or more of these groups with a low-performing subgroup or high-performing subgroup (i.e., “super subgroup”).

Medium: State provides snapshot data of how many students are reaching various achievement levels by subgroup, but does not report either how many students are making expected progress over time or whether the school is improving from year to year. Multiple measures may or may not be reported. Additionally, the state suffers from no more than one of the following: hard-to-find data on the state website; unclear data presentation; or a database that does not permit school-level comparison (though it may permit district-to-district comparisons or school-to-district comparisons).

Weak: State provides snapshot data of how many students are reaching various achievement levels by subgroup, but does not report how many students are making expected progress over time or whether the school is improving from year to year. Multiple measures may or may not be reported. Additionally, the state suffers from two or more of the following: hard-to-find data on the state website; unclear data presentation; or a database that does not permit school-level comparisons (though it may permit district-to-district comparisons or school-to-district comparisons).

- B. State presents data on multiple measures of student performance. These might include attendance, graduation rates, on-track indicators, postsecondary matriculation/retention or remediation rates, course completion (transcript data), attainment of industry certification, and performance on SAT, ACT, and AP/IB exams.

Strong: State presents data on three or more measures of student performance (in addition to proficiency and/or growth data).

Medium: State presents data on at least two other measures of student performance.

Weak: State presents data on one other measure of student performance.

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

- A. Multiple (not fewer than three) user-friendly school and district designations, which rate schools on their effectiveness via accountability provisions, are reported annually (e.g., Florida’s A-to-F system).

Strong: State meets four criteria: 1) It has annual school and district designations in place (three or more for each); 2) those designations are understandable in terms of how they are derived; 3) they include appropriate nomenclature that reflects performance; and 4) each of the labeled categories is structured such that high percentages of schools do not disproportionately fall into the top tiers.

Medium: State has in place at least two of the four above criteria.

Weak: State has in place one or none of the four above criteria.

- B. School and district designations are based, in part, on measures of individual student growth (either normative growth or growth to standard). Further, the designations include performance of specific student groups or a “super subgroup” (e.g., a subgroup combining students by race/ethnicity, English language learner and special-needs status, income, lowest/highest performance, etc.).

Strong: School and district designations are based, in part, on measures of individual student growth. They also include results disaggregated by multiple subgroups or one or more “super subgroups.” The state may also specify whether students are on track to meet graduation requirements and/or college- and career-ready standards.

Medium: School and district designations are based, in part, on cross-sectional growth or school or classroom growth. They also include results disaggregated by multiple subgroups or one or more “super subgroups.”

Weak: School and district designations are based solely on student proficiency (i.e., a fixed measure) or proficiency and other criteria but not growth; and/or they do not include performance of subgroups.

4) A system of rewards and consequences to drive improvement at the school and district levels

- A. Accountability system provides incentives for all schools and districts to achieve at high levels, including, but not limited to, bestowing “honor roll”

or similar “blue ribbon/distinguished” status on deserving schools and districts; offering financial rewards; permitting increased autonomy in operations/spending; allowing other regulatory relief (e.g., automatic renewal of district accreditation) in exchange for meeting heightened district accountability provisions and/or as a reward for performance or growth.

Strong: System provides more than two types of incentives for both schools and districts to achieve at high levels (other than the state accountability designation labels). Incentives at each level must include financial rewards, increased autonomy, or other regulatory relief, and can also include “status” recognition.

Medium: System provides at least one incentive for schools and districts to achieve at high levels (other than the designation labels; can include “status” recognition).

Weak: System provides no incentives for schools or districts to achieve at high levels (other than the state accountability designation labels).

- B. Accountability system requires targeted interventions for low-performing schools, such as replacement of the majority of the staff, charter conversion, state takeover, contracts with outside entities to operate the school, and automatic school closure after consecutive years of failure. (Any or multiple interventions may also be catalyzed via “parent trigger” provisions.) Further, system does not include loopholes by which low performers can avoid tough sanctions; does not water down one or more sanction(s); does not delay sanctions for unreasonable time periods; and/or does not restrict sanctions to fewer than 5 percent of the lowest-performing schools.

Strong: Multiple concrete sanctions exist for low-performing schools, which include automatic school closure after repeated failure; no loopholes or other sanction-avoidance mechanisms in place.

Medium: At least one concrete sanction exists for low-performing schools and loopholes and other sanction-avoidance mechanisms are minimal.

Weak: System provides no meaningful sanctions for low-performing schools, or meaningful sanctions are one option alongside weaker sanctions (i.e., easy loophole).

5) A system of rewards and consequences to drive improvement at the individual student level

- A. State has multiple methods by which students are held to account individually for their performance.

Strong: State has multiple methods by which students are held to account individually for their performance; these may include, but are not limited to: requirements that students pass cumulative high school exit exams in the core subject areas in order to receive a diploma; pass end-of-course exams to get credit for courses or pass end-of-course exams that are integrated into final course grades; pass “gateway” assessments in one or more key grades and subjects in order to be promoted to the next grade (e.g., pass a reading test in elementary school). State may also have a “no pass, no play” policy that requires students to pass all of their courses in order to participate in extracurricular activities, or a “no pass, no driver’s license” policy. States might also reward individual students via guaranteed entry into college and/or guaranteed college assistance (or “priority status”) for eligible students with high GPAs, SAT, or ACT scores.

Medium: States have one or two mechanisms in place to hold individual students accountable for their performance, including one of the following: requirements that students pass cumulative high school exit exams in the core subject areas in order to receive a diploma; or pass end-of-course exams to get credit for courses or end-of-course exams that are integrated into final course grades; or pass “gateway” assessments in one or more key grades and subjects in order to be promoted to the next grade (e.g., pass a reading test in elementary school).

Weak: States have no policies that hold individual students accountable for their performance.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

- A. State requires annual evaluations of all teachers and principals.

Strong: Annual evaluations for all teachers and principals, regardless of experience or status, are required.

Medium: Annual evaluations for some teachers and/or some principals are required.

Weak: Annual evaluations for neither teachers nor principals are required.

- B. State requires that teacher evaluations for all core subject areas include measures of student performance and/or growth on state assessments (assumes that student and teacher data can be linked); state also allows results from the evaluation to be used for continuing employment decisions.

Strong: Teacher evaluations for all core subject areas include measures of student performance and/or growth; state allows results to be used for continuing employment decisions.

Medium: Only in math and reading do teacher evaluations include measures of student performance and/or growth; state allows results to be used for continuing employment decisions.

Weak: Teacher evaluations do not include measures of student performance and/or growth, or student-teacher data cannot be linked; and/or state prohibits results to be used for continuing employment decisions.

- C. State has in place policies that incentivize individual teacher effectiveness and productivity by requiring that districts factor teacher performance into their salary scales (e.g., individual merit pay and/or other individual performance-based bonuses for tasks such as teaching in high-needs schools).

Strong: State requires that districts factor teacher performance into their salary scales.

Medium: State permits/encourages districts to factor teacher performance into their salary scales.

Weak: State prohibits districts from factoring teacher performance into their salary scales.

- D. State has in place other policies that incentivize individual teacher effectiveness and productivity, including, but not limited to: extended contracts, career ladders (e.g., master teachers), personnel decisions based on performance (e.g., no “last in, first out” policies), etc.

Strong: State has two or more of these policies in place to incentivize individual teacher effectiveness.

Medium: State has at least one of these policies in place to incentivize individual teacher effectiveness.

Weak: State has none of these policies in place to incentivize individual teacher effectiveness, or has ill-advised policies that do the opposite (e.g., confer tenure on all teachers in three years or fewer, specify that tenured teachers need not be evaluated annually, etc.).

- E. State mandates that principals be evaluated, in part, according to school performance and that those evaluations inform employment status.

Strong: State mandates that principal evaluations be based, in part, on school performance and allows those results to inform employment status.

Medium: State mandates that principal evaluations be based, in part, on school performance but does not allow those results to inform employment status.

Weak: State does not mandate that principals be evaluated, in part, based on school performance.

- F. State offers incentives for effective principals including bonuses, extended contracts, enhanced autonomy, etc. “Strong” states may also offer incentives for superintendents and/or school board members based on performance.

Strong: State has more than two policies in place that incentivize effective leadership, including merit pay/performance-based bonuses.

Medium: State has at least one policy in place that incentivizes effective leadership.

Weak: State has no policies in place to incentivize effective leadership.

Appendix B: State Accountability Profiles

CALIFORNIA

Overview

California is a leader in establishing robust standards for its students. Unfortunately, the state has not succeeded in translating strong standards into high achievement. Without a strong accountability system to undergird its education system, California standards have fallen flat. The state serves as a primary example of how standards must be paired with equally strong assessments and accountability in order to drive achievement at high levels. While the Golden State has dabbled in a number of different accountability models since its enactment of the Public Schools Accountability Act of 1999, these models—all tied to additional funding, and all optional for schools and districts—have since been abandoned.

Today, California has in place few—if any—mechanisms through which it might hold its schools, districts, students, and teachers accountable. Outside of meeting federal regulations, the state does not tie any mandatory incentives or sanctions to a school or district’s academic performance; it simply reports scores. Teacher and administrator provisions are equally lacking, if not worse: While other states have begun dismantling seniority-based pay scales and layoff provisions, California continues to permit the former and mandate the latter. The only area in which California is modestly successful is in holding students accountable. Students must comply with “no pass, no play” regulations, and pass a high school exit exam in order to graduate (though this exam is primarily based on sixth- and seventh-grade standards).

Few would make the case that California’s state accountability system is now (or ever was) well-equipped to ensure that its strong standards—which now include the Common Core—drive student achievement.

Below, we map California's progress against six key components of strong state accountability systems.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

California is well known for its content-rich and rigorous curriculum standards. Even before adopting the Common Core standards in 2010, the Golden State posted high marks for its standards across all four core content areas.²⁸ In fact, a 2010 review by the Thomas B. Fordham Institute found that, while the Common Core reading and math standards would be a significant improvement over most states' current standards, California's standards were "clearly superior" to the Common Core.

California measures student progress against its California Standards Tests (CSTs), which are aligned to the state's standards.²⁹ The CSTs comprise numerous assessments and are administered to students in the following subjects and grades: English language arts in grades two to eleven; mathematics in grades two to seven, followed by subject-specific end-of-course mathematics tests through eleventh grade; history in eighth grade, followed by subject-specific end-of-course tests through eleventh grade; and science in grades five, eight, and ten, with subject-specific end-of-course assessments in grades nine through eleven. California also requires students to pass the California High School Exit Examination (CAHSEE) to receive a diploma.

Unfortunately, California does not set a high bar of proficiency for these exams. When mapped onto the proficiency scale of the National Assessment of Educational Progress, California's proficiency cut scores reflected NAEP's *basic* score on eighth-grade reading and fourth-grade math, and NAEP's *below basic* score on fourth-grade reading.³⁰

²⁸ In a July 2010 review by the Thomas B. Fordham Institute, California received As for its reading and math standards; the Common Core standards, which California adopted in August 2010, earned grades of B-plus and A-minus for reading and math, respectively. In 2011, California received an A-minus for its U.S. History standards, and in 2012, an A for its science standards. See *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, and *The State of State Science Standards 2012*, at www.standards.educationgadfly.net/.

²⁹ The state also administers the California Modified Assessment (CMA) and the California Alternative Performance Assessment (CAPA). Together with the CSTs, these comprise California's Standardized Testing and Reporting (STAR) program.

³⁰ No data were available for eighth-grade math, as California does not have a general end-of-year assessment in math for this grade. See National Center for Education Statistics, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>.

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

California provides a rich database of information on its department of education website. Users can easily find data on school and district test scores, performance rankings within the accountability system, SAT/ACT and AP participation rates and results, dropout rates, graduation rates, college-going rates, teacher demographics and experience, and so on. Much of the student data are disaggregated by ethnicity, gender, disability, income level, and English-language fluency, and the assessment results are disaggregated by proficiency level: advanced, proficient, basic, below basic, and far below basic.

But while much information is provided, districts report that they do not receive it in a timely manner and thus cannot use it to improve their schools.³¹ In addition, the state data tool does not allow for easy comparisons among schools and districts. Further, users must run separate reports for each indicator listed above; in other words, indicators are rarely listed on the same page for easy comparison within a single school or district. The department of education website does, however, link to an outside website that collects school and district data and allows easy comparisons among and within those entities.

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

California differentiates school and district performance through its Accountability Progress Reporting system, which includes compulsory state reports—Base and

³¹ California has struggled over the last decade and a half to establish a comprehensive, longitudinal database of student, teacher, and school- and district-level performance. The California School Information Services (CSIS) repository, launched in 1997, was never fully funded and only grew to represent 60 percent of all students in the state. Next came the California Longitudinal Pupil Achievement Data System (CALPADS), established by the legislature in 2002, approved in its final form in 2007, and put in place by 2009-10. According to the Data Quality Campaign, California fails to take a number of necessary steps to put CALPADS data to use: it does not link CALPADS to workforce data, does not provide stakeholders with data in an appropriate timeframe, and does not use the data to create individual progress reports, among other actions. Finally, in 2006, California approved the development of the California Longitudinal Teacher Integrated Data Education System (CALTIDES) to integrate existing systems to allow for longitudinal data on teacher characteristics that could potentially be linked to student achievement; but programming for this initiative was cut by Governor Jerry Brown in June 2011. See Marci Kanstoroom and Eric C. Osberg, *A Byte at the Apple: Rethinking Education Data for the Post-NCLB Era* (Washington, D.C.: Thomas B. Fordham Institute, November 2008), <http://www.edexcellence.net/publications/a-byte-at-the-apple.html>; *DQC 2011 State Analysis—California* (Washington, D.C.: Data Quality Campaign, 2011), http://www.dataqualitycampaign.org/files/state_pdfs/CA.pdf; and Shane Goldmacher, “Gov. Jerry Brown signs ‘honest but painful’ budget,” *Los Angeles Times*, July 1, 2010, <http://articles.latimes.com/2010/jul/01/local/la-me-state-budget-20100701>.

Growth Academic Performance Index (API) results—and federally required reports—Adequate Yearly Progress (AYP) and Program Improvement (PI) results.³²

The state's API calculation incorporates student test results on all four subjects included in the CSTs—English language arts, math, history, and science—as well as on the CAHSEE. Both the Base and Growth API scores are represented by a single number, ranging from a low of 200 to a high of 1000. A school's Base API is determined by averaging its students' performance, which is translated according to the following metric:

- Advanced = 1000
- Proficient = 875
- Basic = 700
- Below Basic = 500
- Far Below Basic = 200

The state ranks a school's Base API against the Base APIs of all other schools statewide of the same level (elementary, middle, or high school) and of one hundred other schools of the same type (i.e., schools with similar demographics). Schools are ranked in ten categories of equal size, from one (lowest) to ten (highest). While these ranks are useful in comparing static performance, they are solely based on Base API and do not incorporate growth measures.

Growth API is calculated in exactly the same fashion and with the same indicators as the prior year Base API, but from test results of the current year, to show the amount of growth made by a school.³³ The state has set 800 as the API target for all schools to meet. Schools that fall short of the 800 target are required to establish annual growth targets until the goal is achieved. Each school in California has a unique yearly growth target based on its current performance level, as well as a set

³² A senate bill (S.B. 547) introduced in the California legislature in 2011 aimed to transform API into an Education Quality Index (EQI) and, for the first time, include measures of the preparedness of pupils for postsecondary education and career. (Currently, assessment results comprise the only criteria.) The bill was favorably received by many in the education community: Interviewed state and district representatives alike were critical of California's sole reliance on assessment scores as the basis for accountability in the Golden State and approved of the inclusion of additional measures. The new bill planned to make EQI operational in the 2014-15 school year and include within it a State Assessment Index, a Graduation Rate Index, a College Preparedness Index, and Career Readiness Index. But while the bill was expected to pass, California Governor Jerry Brown vetoed it in early October 2011, maintaining that it did not go far enough toward reducing the accountability system's reliance on quantitative data.

³³ Because California has altered the assessments included in calculating API (additional grades were added in both 2009 and 2010), comparing Base APIs across those years would not accurately measure growth. Thus, each year a Growth API is calculated using the same assessments and grades as used for the Base API of the previous year, and a new Base API is calculated incorporating the new assessment requirements. For example, the 2008 Base API is compared with the 2009 Growth API, using the same assessments in the same grades in both calculations. Then the 2009 Base API is calculated to incorporate the new assessment requirements. The next year, the 2009 Base API is compared with the 2010 Growth API, using the same indicators in each, and a 2010 Base API is calculated to include any new assessments introduced into API in 2010.

of API growth targets for each student group within the school. (Those schools or subgroups already meeting the 800 target must only “maintain 800 or more.”)

Along with API data, California reports federal AYP results. The federal structure sets different proficiency targets than does the state-required API index, but as all states are required to integrate an “additional” indicator into AYP, California also includes API data *within* its AYP calculations. For example, in 2010-11, schools were required to achieve about 68 percent proficiency in both English language arts and math, as well as reach an API score of 710 or a one-point improvement on the API scale, in order to make AYP. But despite this integration, AYP continues to be seen as wholly different from the API-based state system. An education department staffer characterized the perception in this way: “API is an improvement-based system that looks at aggregate-level scores with the inherent focus on moving students upwardly at all levels. AYP only measures the difference between proficient and not proficient.” Given the conflicting proficiency targets—and thus conflicting information on school performance displayed by API and AYP—all parties acknowledged that the current system creates ambiguities for schools and districts.

4) A system of rewards and consequences to drive improvement at the school and district levels

Rewards

The state accountability system in California provides no concrete incentives or rewards for high-performing schools—which is no surprise, given the state’s poor fiscal state of affairs. Instead, the state relies on labels of commendation to incentivize high achievement. The California Distinguished School Award program invites schools that have narrowed achievement gaps and met other API benchmarks to apply for recognition. Selected schools receive a Distinguished School plaque and flag. In addition, national Blue Ribbon awards and NCLB-required Title I Academic Achievement Awards are supposed to drive achievement to high levels.

But as one district representative described, these elements do little to motivate schools and districts to improve above a minimal level of performance: “Sure, it is an incentive not to be labeled low-performing and, sure, we have Blue Ribbons in place, but [these incentives are] not genuine or meaningful.” According to another district-level staffer, “It really comes down to the carrot versus the stick; ultimately, the incentive in California is to keep away from the stick.”

This was not always the case: California has cycled through a number of monetary incentive systems, but each in turn has folded due to budgetary constraints. The California Department of Education recognizes this shortcoming, however, and

assures observers that it is working to implement an incentives program in the future—budget permitting.

Sanctions/Supports

One district representative described California's school and district provisions this way: "Though California's accountability system has been in place a long time and there is a lot of energy behind it...the accountability system is not working well at this point in time." That is a vast understatement. The Golden State has in place few accountability provisions outside of federal NCLB requirements. While it layers its own performance labels—known as Performance Improvement (PI) designations—on top of the federal system, these labels, which are based on the number of years a school or district has missed AYP, are purely descriptive and entail no additional consequences beyond those based on AYP.³⁴

District representatives report that NCLB sanctions have little impact on schools and are unevenly implemented. As one put it, the accountability system is "aggressive about identifying the low-performing [students and schools]" but the federal sanctions "[don't] make a big difference." The district representative also reported that the state provides minimal direction as to which interventions are most important and which parties are responsible for implementing them—and, in its current financial state, the state certainly doesn't step up to the plate. With scant support and direction emanating from the state's central offices, little is done to ensure that improvement efforts are faithfully implemented at the district level.

The PI label, however, does seem to serve one purpose: stigmatization. According to a state representative, "Everyone wants to avoid being in PI." Teachers and principals, worried about their reputations, work to avoid PI status.

Beyond PI requirements, California operates a limited intervention program for low-performing schools. The Quality Education Investment Act (QEIA), the result of a 2006 lawsuit over the state's failure to meet its financial obligations, is an optional one-time, seven-year program that will direct about \$3 million to schools ranking in the bottom 20 percent according to 2005 API scores.³⁵ The program

³⁴ The PI distinction, on its own, is simply a label describing the recent years in which a school has made AYP. Schools and districts "Entering PI" are those that have failed to meet AYP requirements in the same area for two years. Those "Advancing in PI" are schools and districts already in PI that fail AYP again in the current year, while those already in PI that meet AYP in the current year are labeled "Maintaining PI." Any school or district in PI that meets AYP for two years is labeled "Exiting from PI."

³⁵ QEIA is just one in a line of intervention programs that California has introduced and subsequently abandoned. The Immediate Intervention/Underperforming Schools Program (II/USP), part of the original Public Schools Accountability Act of 1999, provided funding for schools in the lower half of API rankings making inadequate growth; in turn, schools were to develop improvement plans, and if they did not meet conditions outlined therein, face potentially serious consequences, such as state takeover. This program was soon replaced by the High Priority Schools Grant Program (HPSGP), established in 2001, which doubled the II/USP funding for individual schools but targeted those funds to the bottom 10 percent of schools. Compared

runs from 2007-08 to 2013-14; participating schools must agree to meet annual API targets, reduce class size, and provide a counselor for every 300 high school students. They also must ensure that their teachers are highly qualified (according to NCLB's highly qualified provision) and that the school's average teacher experience level meets or exceeds that of other like schools in the district. Districts with schools receiving QEIA funds must complete annual reviews for each participating school, ensure administrators are highly qualified, and meet textbook and facilities requirements, among other stipulations. Compared with accountability provisions outlined in other state profiles in this analysis—and given that participation in QEIA is *voluntary* for schools—few could argue that California has implemented a meaningful system that aggressively drives low-performing schools and districts to improve.

One bright—and unique—spot among California's school accountability provisions, however, is recent legislation that allows parents at persistently low-performing schools to organize to bring about radical change. If 51 percent of parents at a school that has missed AYP for six consecutive years (known as schools in "PI Year 4") vote to pull the "parent trigger," they can force the school to replace its principal, replace its staff, or convert to a charter school, among other options. The legislation was hard-fought by Parent Revolution, a group that organizes parents in Los Angeles and throughout California, both before and after its enactment. Once passed, many feared that the state board of education would water down the specific provisions of the law, but parents prevailed in compelling the board to enact regulations upholding the original intent of the law.

5) A system of rewards and consequences to drive improvement at the individual student level

While many of California's school and district accountability requirements have been phased out over the last decade, the state continues to enforce its 1999 mandate that students must pass a high school assessment in order to graduate. The California High School Exit Exam (CAHSEE) assesses students in reading and math and is first administered to students in tenth grade. If students fail, they are able to retake the test two times in eleventh grade and up to five times in twelfth grade. School districts are to provide additional instructional programming for students who continue to fail, either before or after school, during the summer, on Saturdays, or during breaks.

with II/USP, HPSGP required districts to play a greater role in monitoring school improvement and tied more accountability to inputs, such as textbooks and credentialed teachers. HPSGP was superseded by the Quality Education Improvement Act (QEIA) in 2007-08. (Schools participating in HPSGP's final cohort continued to receive funds under the program through 2008-09, but those funds were deemed non-categorical—as in, they were no longer restricted to improvement efforts—in 2008-09.)

Still, the CAHSEE sets an extremely low performance bar, particularly in math. The latter assesses students primarily against sixth- and seventh-grade standards; Algebra I topics comprise less than one-fifth of the overall exam. The reading portion of the exam assesses reading content through the tenth-grade standards—surely a higher bar than that for the math portion, but still relatively low, given the lack of rigor in the state assessment system.

California also imposes a “no pass, no play” policy to drive student achievement. Students in seventh grade and above must maintain a 2.0 grade point average in order to participate in sports and other extracurricular activities.

Finally, California does offer some limited options for students to enroll in schools of choice. Beyond NCLB provisions, the state requires districts to allow intradistrict choice, as long as a chosen school has enough capacity to take on extra students. The state also permits interdistrict choice, though both the sending and the receiving districts must agree to allow it. The number of students partaking in either intra- or interdistrict choice is unclear, as such decisions are made at the local level. California is home to many charter schools, however. About 6 percent of all students were enrolled in charter schools in 2010-11. The state also hosts a number of single- and multi-district virtual schools and programs that students can choose to attend. In 2009-10, about 15,000 students were enrolled in full-time virtual schools, though they comprised less than 1 percent of all students.³⁶

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

California generally takes a hands-off approach to teacher and administrator accountability, letting districts decide how and when to hold accountable adults working in the school system. At the state level, almost no provisions currently exist to incentivize individual high performance or address low performance. The state allows districts to establish their own teacher evaluation systems within a state-determined framework, but the framework is very general: The state does not require that objective measures of student learning be the main criteria included in those evaluations,³⁷ nor does it require that evaluations differentiate

³⁶ John Watson, Amy Murin, Lauren Vashaw, et al., *Keeping Pace with K-12 Online Learning: An Annual Review of Policy and Practice* (Durango, CO: Evergreen Education Group, 2011), <http://kpk12.com/reports/graphics/>.

³⁷ The Stull Act, enacted in 1971, requires that measures of student achievement be included in teacher and administrator evaluations, but it does not stipulate how much weight those criteria should carry. The law was amended in 1999; as it currently reads, educator evaluations must assess teacher impact on student mastery of the district and state standards “as measured by state adopted criterion referenced assessments.” In November 2011, anonymous families in conjunction with Ed Voice, an education advocacy organization, sued the Los Angeles Unified School District (LAUSD) for failing to comply with this regulation. According to Ed Voice CEO Bill Lucia, few of the state’s 1000 districts are in compliance with the Stull Act. See California Education Code 44662 and Larry Sand, “A 40-Year Shame,” *City Journal*, January 19, 2012.

performance across more than *two* ratings. The state only requires that new teachers—those without tenure—be evaluated annually. Tenured teachers may be evaluated every other year, and highly qualified tenured teachers with more than ten years of experience may be evaluated only once every five years. Tenure is (shamefully) conferred virtually automatically after two years.

Teacher evaluations are not tied to employment decisions at the state level. Districts must only offer improvement assistance to teachers earning unsatisfactory ratings, and tenured teachers receiving unsatisfactory ratings must be evaluated annually until they receive satisfactory ratings. The state does not stipulate that teachers with repeated unsatisfactory ratings be eligible for dismissal, or that their tenure status be nullified. Those tenured teachers who are dismissed have multiple opportunities to appeal.³⁸ California also requires that layoffs be based on tenure and seniority; in other words, the state mandates “last in, first out” policies.

California sets a minimum salary but otherwise allows districts to determine their own pay scales. The state does, however, offer select merit pay for high-performing teachers: The Certificated Staff Performance Incentive Act, part of the state’s 1999 accountability legislation, awards funds to high-performing teachers (as determined by the state board of education) in low-performing schools (any school below the 50th percentile). While the state encourages districts to offer additional compensation for teachers working in high-needs schools and subjects, budget woes have eliminated all monetary support from the state in this area. The state used to offer \$20,000 bonuses for board-certified teachers working at least 50 percent of the time in high-needs schools for four consecutive years, but the program is currently suspended for the foreseeable future.³⁹

California has virtually no provisions in place aimed at evaluating and ensuring administrator quality or holding administrators accountable for performance.

What are the strengths and limitations of California’s accountability system?

Strengths

Strong academic standards. California’s academic standards are rated among the best in the nation, routinely earning top marks in analyses of states’ standards.

³⁸ “Unsatisfactory performance” is listed among various other grounds for termination outlined by the state, along with “immoral or unprofessional conduct,” “dishonesty,” and “knowing membership by the employee in the Communist Party.” See California Education Code 44932.

³⁹ “Board-certified” teachers refers to those certified by the National Board for Professional Teaching Standards.

Targeted approach to identifying low performance. While California is not always quick to take effective action in improving low performance, it is aggressive about identifying low-performing schools and student subgroups. District representatives repeatedly referenced the desire not to be identified a school or district in PI.

Limitations

Limited indicators used to differentiate school and district performance. The state relies solely on assessment results to judge school and district performance; no other performance indicators are used. Additional measures such as graduation and dropout rates would help gauge whether schools are ensuring that graduating students are ready for college or careers. In addition, while school and district API scores include aggregate growth, an individual-student growth measure would go a long way toward ensuring that all students are on track for graduation.

Confusion with different proficiency standards and growth measures under federal and state accountability systems. The different performance targets of the state and federal measures, API and AYP, create confusion for parents, community members, and school staff alike. California is one of a handful of states that have opted not to apply for NCLB waivers; this ambiguity will remain unless California specifically acts to remedy it.

Limited system of rewards and sanctions. The Golden State's accountability system does not reward impressive growth or performance. Nor does it impose consequences for low-performers, outside of NCLB. Further, district representatives report that even those NCLB-required actions are not adequately supported by the state. Certainly transparency is a necessity in twenty-first-century state accountability systems, but it cannot be expected to drive school and district improvement by itself.

Limited adult accountability. California has no system in place to ensure highly effective, high-quality staff. The state does not require that measures of student achievement be the chief criteria in teacher evaluations—nor does it appear to hold districts to the requirement that they be included at all—and it does not compel districts to base employment decisions on evaluations. (We're told, too, that the state is not considering adding such policies in the future.) Further, according to district representatives, the state adamantly discourages districts and schools from experimenting with these initiatives on a local level.

Final Word

California operates a subpar accountability system that relies on the ill-functioning federal NCLB system requirements for incentives and sanctions. The state falls short in turning data into action. Without a robust accountability system, the Golden State cannot expect its strong foundation of curricular standards to take root. And the hard work of schools, districts, and students will not be amplified or rewarded.

Information on California's education-accountability system was primarily drawn from interviews with state representatives, district representatives, and local stakeholders, as well as from the California Department of Education website at www.cde.ca.gov. Additional information was drawn from EdSource (www.edsource.org) and from the National Council on Teacher Quality's *2011 State Teacher Policy Yearbook*.

COLORADO

Overview

With the passage of its Education Accountability Act in 2009, Colorado introduced a new accountability system designed to ensure that all students exit the K-12 system ready for postsecondary education and/or the workforce.⁴⁰ The system's theory of action centers on the transparency of its performance-data reporting: With transparent data, the state aims to engage all stakeholders and ensure that information is used purposefully. The 2009 legislation also revamped the state's support and intervention models to provide a more cohesive structure for school improvement efforts.

In 2010, the state took an additional step toward ensuring accountability at all levels by passing a groundbreaking teacher and principal accountability system that dismantled traditional tenure provisions. The state now bases employment on performance measures, which are primarily informed by student assessment data.

Transparent data and school improvement planning are key elements of strong accountability systems, and the Centennial State has placed great emphasis on both. Still, Colorado's schools and students would benefit from improved performance designations that better distinguish among high- and low-performing schools and from more robust interventions for those low performers.

Below, we map Colorado's progress against six key components of strong state accountability systems.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

The Common Core standards in reading and mathematics, combined with the Colorado Academic Standards in all other subjects, result in a motley group of academic expectations that vary dramatically in quality. While the Common Core reading and math standards are considered to be comprehensive and rigorous, Colorado's U.S. History and science standards are poor.⁴¹

Colorado annually assesses students in reading, math, and writing in third through tenth grades, as well as in science in fifth, eighth, and tenth grades. The state is

⁴⁰ The latter is typically defined as the ability to earn a living wage immediately upon graduation.

⁴¹ In a 2010 review by the Thomas B. Fordham Institute, the Common Core standards for reading and math, which Colorado adopted in 2010, earned grades of B-plus and A-minus, respectively. In similar reviews, Colorado's U.S. History standards earned an F (2011), and the state's science standards earned a D (2012). See *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, and *The State of State Science Standards 2012*, at www.standards.educationgadfly.net/.

currently transitioning to a new system that will require annual assessments in reading, math, and writing in third through eleventh grades, as well as in science and social studies at least once in each grade span: elementary, middle, and high school. The new assessment system will also introduce mandatory quarterly assessments, pending available resources. Transitioning to a new assessment system gives Colorado a huge opportunity to improve the rigor of its tests: In a 2011 analysis, Colorado's proficiency cut scores for reading and math in both fourth and eighth grades equated to the *below basic* level of the National Assessment of Educational Progress.⁴²

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Colorado makes a concerted effort to provide transparent and timely data. The state reports data on student proficiency, growth, achievement gaps, and postsecondary readiness in reading, math, and writing through its "SchoolView" website. The online data can be broken out by districts, schools, grades, and subgroups within schools. Further, the Colorado Growth Model, deemed by some stakeholders as the most important feature of SchoolView, allows users to compare the performance of Colorado schools and districts with similar schools and districts to gauge progress.

But while SchoolView provides a wealth of information, it is also difficult to navigate, according to both district and community representatives. Though the state provides materials to aid interpretation, this "guidance" can be so complex as to confuse even the savviest user. At the same time, SchoolView data can be ambiguous. For example, one district representative noted that Asian students, who typically outperform other minority groups, are included in the tool's "minority students" subgroup, thereby skewing the results.

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

Colorado's accountability system measures districts and schools annually against four key performance indicators to inform its accreditation process (the state does

⁴² National Center for Education Statistics, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>. While Colorado is a participating member in both Common Core assessment consortia, it is yet to be seen whether the state will replace its new reading and math assessments with the not-yet-released Common Core assessments (due out in 2014-15).

not include any federal accountability measures, such as AYP, in the accreditation process):

- **Achievement:** the percent of students who are proficient and advanced on state standardized assessments in reading, writing, mathematics, and science.
- **Growth:** normative and criterion-referenced (i.e., growth to a proficiency standard) measures of growth; growth is summarized for a school, district, or subgroup using the median of the student growth percentiles of that group, and then evaluated to determine whether that growth rate is sufficient for the median student in that group to achieve proficiency in three years or fewer.
- **Gaps:** normative and criterion-referenced measures of median student growth, disaggregated by subgroup.
- **Postsecondary and workforce readiness (high school only):** graduation rate, dropout rate, and average composite scores on the ACT.

The Centennial State uses the outcomes from these four broad performance indicators to arrive at overall determinations of school and district performance. At the **district level**, the indicators inform the annual accreditation process. Districts are rated on a five-level scale:

- Accredited with Distinction
- Accredited
- Accredited with Improvement Plan
- Accredited with Priority Improvement Plan
- Accredited with Turnaround Plan

At the **school-level**, ratings are a tad more complicated. The state recommends a preliminary designation of school accreditation on a separate, four-level performance scale. Then, districts—which officially accredit schools—make the final school-accreditation designation. Districts may concur with the preliminary state recommendations, or they can choose to be more stringent in their accreditation determinations. (To be more lenient, a district needs special approval from the state.)

This process gives both the state and districts the opportunity to have a voice in schools' performance designations, which helps balance local and state control. The state engages in light monitoring of districts throughout the school rating process (particularly failing districts), but most parties agree that the ratings are fair; plus, schools can challenge their ratings through an appeals process.

The four-level school-accreditation scale is as follows:

- Performance
- Improvement
- Priority Improvement
- Turnaround

Unfortunately, these levels fail to differentiate school performance adequately. The lowest-performing category, Turnaround, comprises about 5 percent of schools; the next category, Priority Improvement, includes about 10 percent; and the third category, Improvement, generally encompasses 20 percent. This leaves a full 60 percent of schools that fall into the highest category, Performance. In other words, the categories don't differentiate between a school at the 42nd percentile and a school at the 98th percentile. Worse, a local stakeholder reports that parents are confused by the meaning of the names—is Performance better than Improvement? What does Priority Improvement signify? Such ambiguity erodes the value of assigning performance designations in the first place.

4) A system of rewards and consequences to drive improvement at the school and district levels

Rewards

By allowing the top 60 percent of its schools to fall into the Performance category, Colorado offers little ability—much less concrete incentive—for schools to strive to achieve at high levels. District designations, which comprise five categories instead of four, provide slightly more accurate pictures of performance, but this difference is not significant: District representatives made clear that the state does not meaningfully distinguish among categories. The state does honor schools with low-performing students making significant gains with “Center of Excellence” status, but such recognition bestows no meaningful reward. Many district representatives expressed desire for more robust and tangible incentives from the state.

Sanctions

Colorado requires all schools and districts to engage in improvement planning, with the most intensive actions required of the lowest performers. District and school accreditation levels inform the type and depth of required improvement plans. These plans provide information on the district/school's data trends, set achievement targets, and identify strategies and resources the district/school will implement to achieve those goals.

Districts whose performance places them in the top three accreditation categories in a given year have their accreditation automatically renewed with the submission of a Performance or Improvement Plan. Accreditation for the bottom two categories is contingent upon state acceptance of a district Priority Improvement Plan or Turnaround Plan. All districts are expected to use the district-level Unified Improvement Plan (UIP) as a template for these improvement strategies. The template streamlines state and federal accountability requirements into a single, model improvement plan.⁴³

At the school level, the principal, superintendent, and/or local school board, in collaboration with a School Accountability Committee,⁴⁴ work together to develop the improvement plan. Districts are expected to review all school plans, while the state only reviews school Turnaround Plans (and *may* review school Priority Improvement Plans). School accountability committees are required to meet quarterly to review progress toward the goals outlined in the improvement plans.

The bottom two school accreditation categories—Priority Improvement and Turnaround—are considered schools “identified for improvement.” Once identified, a school has five years to improve its accreditation status and reach the goals set forth in its improvement plan. If it fails to do so, the school must restructure or close. (Since these requirements were established in 2009, no school has yet hit the five-year failing mark.)

The school improvement model has some flaws in practice. For one, it limits real sanctions—such as restructure or closure—to those schools that persist in the worst performance categories for five years. Other low-performing—but slightly better—schools go unnoticed. For these other schools, the system is intended to provide “more of a political sanction,” as one district representative put it. In other words, transparent data are meant to catalyze improvement—at least in theory—as parents, students, and community members can now identify which schools need improvement. But as mentioned earlier, the performance data provided by the state are often confusing to stakeholders and fail to distinguish among average and above average schools.

At the same time, state capacity for driving improvement in the lowest-performing schools is in question. Colorado received a \$40 million federal School

⁴³ While Colorado’s school and district performance framework does not incorporate any measures of the federal accountability system, the requirement that schools and districts must address both systems in their UIPs is a step toward easing the burden on schools and districts struggling to adhere to both systems at once. By incorporating both into improvement planning, schools and districts can implement strategies that address state and federal requirements at the same time, rather than operating dual, disjointed, and burdensome improvement programs in parallel.

⁴⁴ The state requires each school to convene a School Accountability Committee, which must comprise, but is not limited to: the principal (or the principal’s designee); a teacher; three parents; an adult member of an organization of parents, teachers, and students recognized by the school; and a community member.

Improvement Grant in 2010 to turn around nineteen of its schools. A year later, only one of the nineteen schools has shown improvement across the board in reading, math, writing, and science. While turnaround efforts often take more than one year to prove successful—and require more than one year to measure that success reliably—the minimal improvements raise flags about whether the state has an effective strategy in place for turning around its lowest performers. Under the accountability system established in 2009, the state must now restructure or close its lowest-performing schools, as described above. Again, the first schools have not yet hit the five-year failing mark; when they do, the state will no longer have the additional infusion of federal turnaround money.

Support

While real improvement under the new accountability structure is yet to be seen, district representatives cite as a positive step the transformation of the state department of education from a regulatory to supportive body. Colorado's method of distributing and offering support to its districts, which are primarily tasked with providing direct assistance to schools, is based on school/district performance and need, as demonstrated through accreditation ratings and improvement plans, as well as AYP measures. The state operates a tiered system of support with three categories commensurate with the level of need, resources, and student performance of schools and districts. Most districts are in the "universal" tier, which provides a baseline level of state support. The state offers some additional support for the schools and districts in the second, "targeted" tier. The highest level of state support is reserved for those in the third, "intensive" tier. This most intensive support is aimed at turnaround schools and includes an on-site team to assist in analyzing data, identifying areas of weakness, and determining priorities, among other elements.

5) A system of rewards and consequences to drive improvement at the individual student level

Colorado's accountability system includes virtually no measures related to individual student accountability. Students are not required to pass a cumulative high school exam in order to graduate, nor are they required to pass cumulative assessments in key grades in order to be promoted to the next grade.

The Centennial State is unique, however, in allowing its students to attend any school in any district, as long as space is available. So while the state doesn't specifically *require* that schools allow students in low-performing schools the option of attending other schools, any student may choose to do so.⁴⁵ Of course,

⁴⁵ As long as that school is public, of course. Colorado does not have a statewide voucher program in place, though districts are free to introduce such options.

policies like these often look good on paper but are not exercised in practice. Moreover, they are often not supported—and may actually be spurned—by would-be receiving districts and schools.

In addition, the state also provides options for students through charters and virtual schools. In 2011-12, about 114,000 students enrolled in Colorado charter schools, comprising about 14 percent of all students. Further, Colorado runs a state virtual school and also permits district-level digital programs; in 2011-12 about 16,000 students registered for online education programs—about 1.9 percent of all students.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

In 2010, Colorado passed S.B. 10-191—the Great Teachers and Leaders bill—which completely revamped the state’s legislation on teacher and principal accountability. It was hailed as a huge victory for Colorado’s students, and by some, as a model for the rest of the country. While the new evaluation system will not be fully implemented until the 2014-15 school year, it will base at least 50 percent of annual teacher and principal evaluation on student achievement data.

Teachers and principals will be rated as highly effective, effective, partially effective, and ineffective. New, or probationary, teachers must earn three consecutive effective ratings to merit the equivalent of tenure; a teacher who earns two consecutive ineffective ratings is placed back into probationary status and is given a year to improve or face dismissal. Districts are given leeway in how to account for low-performing principals.

In addition, districts are required to develop systems for rewarding high-performing teachers who choose to teach in low-performing schools. They may also, but are not required to, introduce merit-based pay systems for teachers and principals.

What are the strengths and limitations of Colorado’s accountability system?

Strengths

Focus on transparent data reporting. Colorado’s SchoolView accountability system releases annual performance data in a timely manner. The database allows users to disaggregate proficiency and growth data by district, school, individual grade, and subgroup. Still, the system leaves much room for improvement, as the information can be misleading and/or confusing.

Comprehensive evaluation of student performance. Colorado's accountability system provides a comprehensive picture of academic performance through its measurement of achievement, growth, gaps, and postsecondary and workforce readiness.

Responsibilities distributed among stakeholders at all levels. Under Colorado's accountability system, both the state and districts have a say in school performance designations and required improvement strategies. This promotes shared ownership of the problem as well as shared victory in success. In addition, a variety of local stakeholders participate in school improvement planning.

Robust, annual evaluation of teacher and principal performance. Colorado requires that 50 percent of annual teacher and principal evaluations be based on student performance measures. For teachers, both tenure and dismissal are conditional, based on the annual evaluations.

Limitations

System lacks a strategy for identifying high achievement. While Colorado excels in tracking student performance, both in terms of proficiency and growth, the state offers no recognition or designation for achievement higher than "performance," which is already a blurry moniker. Colorado's performance designations could benefit from more clarity and the implementation of a performance rating that recognizes excellence.

Limited system of rewards and incentives for schools. Colorado officials believe that the state's transparent data-reporting system is an incentive and sanction in and of itself, meaning that specific and tangible rewards are not incorporated into the accountability system. At the same time, the state only requires intensive interventions in its lowest-performing schools after five years; even then, its capacity to undertake those interventions is dubious.

Absence of student-level accountability measures. Colorado maintains no student-level accountability measures, such as high school graduation requirements or grade-promotion conditions.

Final Word

Colorado's accountability misses the mark somewhat: While it aims to use transparent data and improvement planning to drive reform, its data can be confusing to stakeholders and its improvement planning requirements are relatively toothless. Colorado's system would benefit from recognizing high

achievement as well as low achievement, and by implementing more meaningful sanctions and incentives to drive school, district, and student performance.

Information on Colorado's education-accountability system was primarily drawn from interviews with state representatives, district representatives, and local stakeholders, as well as from the Colorado Department of Education website at www.cde.state.co.us, and the department's SchoolView website at www.schoolview.org. Additional information was drawn from the National Council on Teacher Quality's *2011 State Teacher Policy Yearbook*.

FLORIDA

Overview

In 1996, Florida’s legislature charged the state board of education with implementing an accountability system and intervening in low-performing schools. In the decade and a half since, Florida’s widely-lauded accountability system has continued to evolve as governors and other state officials have worked to identify more effective accountability measures—and as accountability has expanded to include both students and adults in the system. This is a state that, according to an education department staffer, was “focused on accountability far before [it received] a national spotlight or attention from other states.”

One element that sets Florida’s accountability system apart from the rest is its focus on the progress of students across the achievement spectrum. Their performance is gauged through both proficiency and individual-growth measures; in addition, the Sunshine State tracks the progress of the lowest-performing quartile of students in each school. At the same time, Florida continues to re-evaluate the system’s expectations and raise the bar for performance, maintaining tension in the system. Its focus on academic improvement for all students and its willingness to keep pressure on the system may be behind the state’s recent successes, including its progress in shrinking both “achievement gaps” and “opportunity gaps.”⁴⁶

Florida also ensures that individuals are held to account and are rewarded for strong performance. Students must meet promotion and high school graduation requirements, and the top 20 percent of each graduating class is eligible for automatic admission into one of Florida’s eleven state universities. Teachers in Florida’s school system are also accountable for their performance and that of their pupils via annual evaluations tied to student learning, the elimination of tenure, and layoffs based primarily on evaluation results, not seniority.

Florida’s focus on accountability is palpable throughout the system. The Sunshine State’s governing bodies are intent on ensuring that education remains a top priority on the political agenda: Candidates for state office are implicitly expected

⁴⁶ See National Center for Education Statistics, *Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress* (Washington, D.C.: U.S. Department of Education, June 2011), <http://nces.ed.gov/nationsreportcard/pdf/studies/2011459.pdf>; National Center for Education Statistics, *Achievement Gaps: How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress* (Washington, D.C.: U.S. Department of Education, July 2009), <http://nces.ed.gov/nationsreportcard/pdf/studies/2009455.pdf>; and “The Opportunity Gap: Is Your State Providing Equal Access to Education? Educational Access in Florida,” ProPublica, June 2011, <http://projects.propublica.org/schools/states/fl>.

to make education a top campaign priority in order to have any chance of being elected.

Below, we map Florida's progress against six key components of strong state accountability systems.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

With its adoption of the Common Core, Florida has in place strong state standards in reading and mathematics. Unfortunately, the state's standards in U.S. History and science lack the content and rigor needed to ensure quality instruction in those subjects.⁴⁷

Since 1975, Florida has required annual testing of all students in select grades and subjects. Today the state's assessment system, the Florida Comprehensive Assessment Test (FCAT), mandates continuous testing in grades three through ten in both reading and mathematics, as well as science in fifth, eighth, and eleventh grades, and writing in fourth, seventh, and tenth grades.

Florida is in the process of transitioning to the FCAT 2.0, which is aligned to the state's latest installment of state standards and will come online in 2011-12. The new assessment system will introduce end-of-course tests in high school for math, science, and history courses, in addition to high school-level reading and writing assessments. (Florida will likely replace the FCAT 2.0 for reading and math with Common Core assessments in the future.) In anticipation of the transition, the state board of education voted in December 2011 to raise the assessments' proficiency cut scores for reading and math at every grade—for the first time in a decade.⁴⁸ Pressure to "raise the bar" is a common thread that runs throughout the Florida accountability system.

⁴⁷ In an analysis conducted by the Thomas B. Fordham Institute in 2010, the Common Core reading and math standards earned grades of B-plus and A-minus, respectively. In similar analyses, Florida's U.S. History standards earned a C, as did its science standards. See *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, and *The State of State Science Standards 2012*, at www.standards.educationgadfly.net/.

⁴⁸ Raising proficiency cut scores is a strong step forward for Florida. In a 2011 analysis, the state's cut scores for reading and math for both fourth and eighth grades were found to equate to the *basic* level on the National Assessment of Educational Progress. See National Center for Education Statistics, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>.

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Florida presents a variety of data on its state department of education website. Users can break out FCAT results by district, school, and subgroup, as well as by five proficiency levels. An easy-to-navigate tool allows users to populate and compare results for multiple schools at once. In addition, the state provides a “performance report” tool that provides data on proficiency levels in addition to school performance designations, graduation and dropout rates, AYP, and more. However, this tool only allows users to view data by legislative district, a curious and relatively unhelpful choice.

For those who are willing to dig, the state does provide data on its website for a number of other indicators, such as attendance and non-promotion rates, staff experience and demographics, enrollment totals, SAT/ACT score averages, and more. Whether or not the data are disaggregated by district, school, or subgroup depends on the particular indicator.

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

Under the leadership of former Governor Jeb Bush in 1999, Florida implemented a school and district grading system, the *A+ Plan*, to which many attribute the Sunshine State’s impressive progress in recent years. District representatives report that the accountability system enjoys widespread support due to its long-standing credibility.

The *A+ Plan* is comprised of four components: standards, assessment, progress reporting, and consequences. Undergirding the structure is an intuitive A-to-F grading scale for schools and districts, which bases final grades on both proficiency and individual-student growth measures. (Proficiency is measured against FCAT scores in reading, math, writing, and science, while growth is measured against reading and math.⁴⁹) Individual-student growth is calculated both for all students and for the lowest-performing quartile. At the high school level, the A-TO-F grades also incorporate other factors such as graduation rates (both for an entire student body and for a school’s at-risk students) and performance and participation rates in AP, IB, dual enrollment, Advanced International Certification of Education, and industry certification courses.

⁴⁹ A state representative reported that Florida will likely decrease the influence of writing scores as it transitions to the Common Core standards, partly because the writing results are more prone to error.

School and district grades are translated as follows:

- A = Excellent progress
- B = Average progress
- C = Satisfactory progress
- D = Less than satisfactory progress
- F = Failing to make adequate progress

Marks like these set the standard for easy-to-understand grading systems across the nation. Still, as explained below, the structure of consequences and supports that builds off of the grading system is not equally straightforward. In addition, because the grading system does not incorporate measures of AYP, schools and districts can score high on the state scale and low on AYP, or vice versa. As one district representative put it, “Parents remain confused by a system that can rate a school well and poorly at the same time.”

Florida has set an example, however, for continually setting higher expectations for student achievement. In May 2011, the state board of education passed “automatic trigger” regulations mandating that, in the case that 75 percent or more of all schools earn A or B grades, the scores required to earn those grades must automatically be increased. The state has also raised the cut score for its writing assessment twice since *A+ Plus* began. As one stakeholder put it, this constant pressure increases student achievement: “If we had just kept our original accountability system in place and not ‘raised the bar’ on expectations or other criteria, we would have plateaued in performance long ago.”

4) A system of rewards and consequences to drive improvement at the school and district levels

Rewards

Florida's incentives program is designed, according to a state staffer, “to incentivize high or improved performance as much as it is designed to differentiate between low- and high-performing schools.” Through the Florida School Recognition Program, schools earning A grades, as well as those that improve a letter grade from one year to the next, receive grants totaling \$75 per full-time student. Schools are required to use their awards for any or all of the following: (1) awarding faculty and staff bonuses; (2) purchasing educational equipment and materials; and (3) hiring temporary personnel to assist in maintaining or improving student performance. If the school doesn't use the money within a certain timeframe, the funds automatically go toward teacher bonuses. Still, district representatives weren't ecstatic about these incentives, relaying that incentives provide little motivation beyond the existing expectation that schools

should be high-performing. One district representative noted that “school grades and AYP proficiency are strong drivers at the district level,” not the incentives.

Sanctions

The state uses both school grades and AYP results to determine the degree of technical assistance and intervention provided to its lowest-performing schools. This combination of accountability systems to distinguish between low- and high-need schools is known as Florida’s *Differentiated Accountability* model. The state signed on to a federal pilot project to implement the model in 2008; in brief, it allows the state some flexibility in aligning state and federal accountability systems to reduce redundancy and confusion. The state and federal structures continue to operate independently of each other, but the model allows the state to streamline what schools and district are asked to do under both systems so that they are implementing cohesive improvement strategies—not overlapping or conflicting ones.

Schools not required to participate in *any* improvement strategies are those elementary and middle schools that receive A, B, or C grades and those high schools with FCAT performance scores of 435 or higher; in addition, these schools must have made AYP for at least two consecutive years. All other schools are placed into one of the following categories:

- Prevent I
- Correct I
- Prevent II
- Correct II
- Intervene

These designations are not nearly as intuitive as Florida’s A-to-F performance scale. One district representative reports that “the public pays more attention to school grades” and that “parents don’t really understand the distinctions under *Differentiated Accountability*.”

The Intervene category entails the most state intervention and support. While the criteria used to determine the different classifications are complex, a school generally enters Intervene status if it receives repeated D or F grades and its percentage of proficient students in reading or math diminishes over five years. Intervene schools must implement one of four options, selected by the district, during the following school year: (1) convert school to a district turnaround school; (2) reassign students and monitor the progress of the reassigned students; (3) close

and reopen as a charter school;⁵⁰ (4) contract with an outside entity to run the school.⁵¹

Support

Florida's system does not allow the state to manage or take over any public schools, but the state and districts share in monitoring schools and providing on-site supports. The specific actions that schools and districts undertake together to improve performance vary with the *Differentiated Accountability* categorizations, but can include school improvement planning, evaluating and replacing ineffective administrators, evaluating a school's instructional programming, and monitoring the progress of improvement efforts.

District representatives report that state supports for low-performing schools are often lacking. Despite following state education department protocol to improve and restructure schools, one district representative described the funds and resources that trickle down from the state as "very limited." Districts also face a dearth of available talent. As another representative put it, "Even if we wanted to remove principals from low-performing schools, the pool of administrators is low, so it is difficult to bring in new principals or other staff members."

5) A system of rewards and consequences to drive improvement at the individual student level

Through its collection of student-centered policies, Florida places responsibility for student achievement on the shoulders of students themselves. Third graders in Florida must earn a Level 2 or higher on the state reading assessment in order to be promoted to fourth grade. ("Level 3" is considered "proficient," or "on grade level.") Those students who fail to do so can also be promoted by demonstrating an equivalent level of performance on an alternative, state-approved test, or by demonstrating adequate reading mastery through a student-performance portfolio.

In addition, Florida high school students are required to pass state assessments in order to receive high school diplomas. More specifically, as the FCAT 2.0 is phased

⁵⁰ A 2010 Fordham Institute study found that in Florida, 7 percent of all low-performing district schools and 23 percent of all low-performing charter schools closed between 2003-04 and 2008-09. The study did not report whether these closures were due to low performance or for other reasons. See David Stuit, *Are Bad Schools Immortal? Turnaround and Shutdowns in Both Charter and District Sectors* (Washington, D.C.: Thomas B. Fordham Institute, 2010), <http://www.edexcellence.net/publications-issues/publications/are-bad-schools-immortal.html>.

⁵¹ Schools that persistently fail AYP, but achieve A, B, or C grades under the Florida grading system, are placed in the category above Intervene (known as "Correct II"). So in practice, while the options outlined for Intervene schools parallel NCLB sanctions, schools that repeatedly miss AYP but earn strong grades under state accountability measures are *not* required to implement these most intrusive interventions.

in, students will have to pass the FCAT 2.0 tenth-grade reading test (beginning in 2010-11), the end-of-course Algebra I test (beginning in 2011-12), and end-of-course tests in biology and geometry (beginning in 2012-13). Students can retake the tests as many times as they like until they pass; they can also enroll for a free thirteenth year of public education if they need additional instructional time to pass the assessments. In addition, students can graduate by receiving concordant scores on the ACT or SAT, as determined by the state. The required scores for reading (verbal) increased between 2009-10 and 2010-11, while required scores for mathematics decreased somewhat.⁵²

One stakeholder remarked that the state faces a political disadvantage in mandating that students pass assessments for high school graduation: Florida sees fewer students graduate high school than it otherwise would, and thus doesn't post graduation rates as high as some other states. The state's commitment to student accountability, however, overshadows such concerns.

As an incentive to strive for high performance, Florida high school students can earn automatic admission into one of Florida's eleven state universities if the student is in the top 20 percent of his or her class and completes all graduation requirements and core course requirements for state university admission. These students are considered a priority for the awarding of funds from the need-based Florida Student Assistance Grant.

Beyond incentives and consequences, Florida provides extensive options to support individual low-income students and students in low-performing schools—far more options than offered by most other states. The Opportunity Scholarship Program (OSP) offers students in low-performing schools the option of transferring to a higher-performing school in the same district or in another district (space permitting).⁵³ Like many other states, Florida also offers school choice through its many charter and magnet schools, but it is one of only a handful of states that offer the option of full- or part-time virtual schooling to every child in the state through a state-level online program, the Florida Virtual School (FLVS).

And in Florida, options extend beyond public schools alone: The Florida Tax Credit Scholarship Program (FTCS), established in 2001, provides tax credits for organizations to encourage them to support nonprofits that offer private school scholarships for low-income students. In 2010-11, almost 35,000 students enrolled

⁵² Required scores for the ACT and SAT reading assessments have grown from 410 to 420, and from 15 to 18, respectively. The required score for the SAT mathematics assessments has declined from 370 to 340; that for the ACT has remained the same at 15.

⁵³ Low-performing schools are defined as those with school grades of D or F under the *A+ Plan*, and those in either Correct II or Intervene under *Differentiated Accountability*. At its inception in 1999, OSP permitted students to transfer to participating private schools as well, but this option was declared unconstitutional by the Florida Supreme Court in 2006.

in over 1,000 participating private schools through the program. Similarly, the McKay Scholarship program offers private school vouchers to students with disabilities.⁵⁴

Finally, Florida also requires that districts adopt “controlled open enrollment” policies, which must include parental preference as a large determining factor in student assignments. Taken together, Florida makes plain that it expects its students and parents to take active roles in schooling, both by striving to meet standards, and by choosing the educational environments that best meet student needs.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

In Florida, teacher and administrator accountability are now key parts of the state’s accountability strategy. With the passage of the Student Success Act in spring 2011, Florida ended teacher tenure, established performance pay, and made student learning—as measured by state assessments—the chief criterion in teacher and principal evaluations.

Under the new system, districts may design their own teacher evaluation and compensation systems, but these systems must adhere to state criteria and be approved by the state. Teachers must be evaluated annually (and new teachers, biannually) against four performance categories: highly effective, effective, needs improvement, and unsatisfactory. Tenure decisions—i.e., moving from probationary to non-probationary status—are to be informed by evaluations, not seniority. However, since the state now also mandates one-year contracts for teachers, tenure is essentially eliminated. Teachers rated as “unsatisfactory” must be placed on an improvement plan; those who fail to meet the plan’s outlined criteria are eligible for dismissal. Finally, Florida now requires that layoff decisions be based on evidence of student performance, not on seniority.

Districts must also now base salary schedules on teacher performance, not on seniority (beginning in 2014). So, while districts may design their own salary scales, these must ensure that only highly effective and effective teachers are eligible for salary increases. Teachers can also earn additional compensation by teaching certain subjects or teaching in Title I schools or Correct II or Intervene schools.

⁵⁴ All together—including students participating in OSP, FLVS, FTCS, and McKay Scholarships, and those enrolled in charter schools—Florida saw about 473,000 students take advantage of choice options in 2010-11, or about 18 percent of the total student body. (Author’s calculation, based on enrollment trends drawn from the Florida Department of Education website. Note that this figure may be slightly inflated, as some students that participate both in charter schools and in FLVS, for example, may be counted twice.)

As at-will employees, Florida's principals have no tenure rights. As one stakeholder described, this strengthens the accountability surrounding not just principals, but superintendents as well: "Since Florida grades schools in a clear manner (A-to-F), there is great pressure on superintendents to make wise decisions on principal placements and selection since the grade of the school reflects on the entire district." The state now mandates annual evaluations of principals and requires that student achievement be the main criterion in those evaluations. Principals, like teachers, are only eligible for salary increases if they are rated as highly effective or effective.

What are the strengths and limitations of Florida's accountability system?

Strengths

Comprehensive evaluation of school performance. Florida provides a comprehensive look at academic performance by evaluating schools on the basis of both proficiency and individual student growth. The state's focus on student achievement rates, student learning gains in mathematics and reading, and improvement for the lowest quartile of students is superior to most states' focus on proficiency rates alone.

Emphasis on improving the performance of low-achieving students. As indicated above, an important strength of Florida's accountability system is its focus on the performance and growth of the lowest-performing quartile of students at every school.

Strong system of sanctions and incentives to drive school and district performance. Florida's accountability system provides concrete incentives for schools to perform at high levels, and consequences for schools that perform poorly. While the efficacy of these components is questioned by some districts, others describe them as being "incredible drivers" for improving school and student performance.

Strong focus on student accountability. Florida sets standards for individual students, incentivizes them to meet those standards, and ties consequences to falling short of those expectations. In addition, students benefit from numerous educational options.

Exemplary teacher and principal accountability policies. With the passage of the Student Success Act in spring 2011, Florida ushered in a new era of individual adult accountability in its schools. By requiring annual evaluations, linking them to objective measures of student learning, and eliminating teacher tenure, Florida has ensured that teachers and principals

are responsible for their own performance—and face consequences if they do not perform up to par.

Limitations

Confusing school and district accountability categories. As straightforward as Florida’s A-to-F rating system is, the state introduced confusion into its accountability system by combining the A-to-F ratings with AYP measures to determine school and district intervention categories—and then exacerbated the problem by naming those categories Prevent I or II, Correct I or II, and Intervene. Florida would do well to establish intervention categories that are as straightforward and comprehensible as its A-to-F performance categories.

Final Word

Florida’s accountability system can boast a number of strong features: measurement of both student proficiency and growth; a focus on the lowest-performing students; a system that includes both sanctions and rewards for low- and high-performing schools; and strong student, teacher, and principal accountability measures. In addition, Florida continually raises the bar for student expectations, so as to avoid plateaus in performance.

Information on Florida’s education-accountability system was primarily drawn from interviews with state representatives, district representatives, and local stakeholders, as well as from the Florida Department of Education website at www.fldoe.org and the website’s school choice sister site at www.floridaschoolchoice.org. Additional information was drawn from the Florida Virtual School website at www.flvs.net and the National Council on Teacher Quality’s 2011 *State Teacher Policy Yearbook*.

INDIANA

Overview

Passed in 1999, Public Law 221 (P.L. 221) established Indiana's accountability system for K-12 education. The Hoosier State has continued to tweak and improve the system for over a decade since, most recently through 2010 legislation that restructured the accountability framework to adhere to an A-to-F grading scale for school classification (similar to Florida's system), among other changes.

Indiana premises its accountability system on the notion that, by focusing on progress and providing educators and administrators with diagnostic and monitoring tools, it can induce schools to improve learning outcomes for students. To that end, the state education department provides school districts with a number of optional diagnostic and formative assessments to measure student progress.

At the same time, the state's focus on using data to drive improvement means that only the very worst schools face any consequences and interventions—and only after many years. Interventions exist only for schools that persistently rank in the lowest performance category; no improvement actions are required of any other schools, and no accountability measures are required of districts. This may be a reasonable course of action for Indiana—after all, state education departments are limited in their capacity to intervene and intervene *well*—but it is significant tradeoff nonetheless.

Below, we map Indiana's progress against six key components of strong state accountability systems.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

Indiana recently adopted the Common Core standards in reading and math and uses the Indiana Academic Standards in all other subjects. The standards are considered well developed and rigorous.⁵⁵ To measure student progress, the Hoosier state administers the Indiana Statewide Testing for Educational Progress Plus (ISTEP+) assessments in reading and math in grades three through eight, as well as science in grades four and six, and social studies in grades five and seven.

⁵⁵ In reviews by the Thomas B. Fordham Institute, Indiana received an A-minus for its U.S. History standards in 2011 and an A-minus for its science standards in 2012. The Common Core standards for reading and math, which Indiana adopted in 2010, earn grades of B-plus and A-minus, respectively. See *The State of State U.S. History Standards 2011*, *The State of State Science Standards 2012*, and *The State of State Standards—and the Common Core—in 2010*, at www.standards.educationgadfly.net/.

ISTEP+ also includes high school end-of-course assessments for students completing Algebra I, Biology I, or English 10.

Indiana would do well to improve the rigor of its assessments. While many states establish lower cut scores of proficiency, Indiana's cut scores for eighth-grade reading and fourth- and eighth-grade math all equate to the *basic* level on the National Assessment of Educational Progress (NAEP); worse, its fourth-grade reading cut score equates to NAEP's *below basic* level.⁵⁶

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Indiana fares moderately well on this component. The state presents information on school and district performance through an easy-to-navigate database called the DOE Compass. The database includes information on proficiency rates, graduation rates, dropout rates, AP/IB passing rates, SAT/ACT scores, and AYP data. Unfortunately, the state does not disaggregate any data by subgroup (outside of the AYP data, which only include subgroup proficiency rates), and it only displays student performance by proficiency rate instead of also by performance level.

The state does, however, allow parents direct online access to their children's data, including longitudinal results. Further, Indiana reports a measure of growth known as the Indiana Growth Model. The growth model evaluates individual-student growth relative to peers with similar academic histories, and classifies student growth into three bands:

- **High Growth** – 66th to the 99th percentile
- **Typical Growth** – 35th to the 65th percentile
- **Low Growth** – 1st to the 34th percentile

For school and district reporting purposes, the median student growth percentile in a school or district represents that entity's growth.

⁵⁶ National Center for Education Statistics, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>. For reading and math, the state plans to participate in new Common Core assessments beginning in the 2014-15 school year, as it is a governing state in the Partnership for Assessment of Readiness for College and Careers (PARCC).

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

Indiana's state board of education approved a new metric in November 2011 by which the state will now evaluate schools and districts. While Indiana's previous metric included both state-required and federal AYP measures, the new metric exorcizes all federal measures.

Elementary and middle schools evaluations are now based on proficiency, growth, and assessment participation rates. Schools receive preliminary scores based on their proficiency rates; those scores can be raised if the lowest performing quartile or the remaining 75 percent of students show high growth, as measured by the Indiana growth model. The preliminary scores can also be lowered if a certain percentage of all students show low growth, or if assessment participation rates are low.

High schools are evaluated based on student performance on end-of-course assessments in English 10 and Algebra I, graduation rates, and post-secondary readiness indicators—including AP/IB passing rates, college credit attainment, and industry certification. All three areas are added together to form a final score. Schools receive preliminary assessment scores based on the percentage of students passing the end-of-course assessments; those scores can be raised or lowered depending on student progress from eighth to tenth grade, and from tenth grade to graduation. Schools also receive preliminary graduation scores based on four-year graduation rates—which can be raised or lowered depending on the number of honors and waiver diplomas⁵⁷ issued by the school, and by five-year graduation rates.

Based on these measures, Indiana assigns school ratings on an A-to-F scale:

- A (Exemplary Progress)
- B (Commendable Progress)
- C (Academic Progress)
- D (Academic Watch – Priority)
- F (Academic Probation – High Priority)

Districts are also scored according to the same metric, and are categorized by the same five performance designations.

⁵⁷ A student who does not meet graduation exam requirements may graduate with a waiver diploma if he passes all required courses, demonstrates to the satisfaction of his school that he has met the performance bar set by the state graduation exam, or completes an internship and workforce readiness assessment.

Indiana's removal of AYP in its accountability metric means that traditional subgroups of students are no longer incorporated into the accountability system. As one state representative explained, by focusing on the lowest-performing quartile of students, the state aims to target low performers regardless of race or disability.

4) A system of rewards and consequences to drive improvement at the school and district levels

Rewards

District representatives indicate that the greatest incentive for school improvement is the state's designation system and the accountability requirements tied to different labels. Beyond the school ranking system, Indiana has two incentive programs of note, one of which entails financial rewards:

- **Four Star Schools Awards** – The Four Star Schools Award recognizes schools making AYP that also rank in the state's top 25 percent of schools based on ISTEP+ results in reading and mathematics in all grades tested. Winning schools are recognized through various press releases and are featured on the state department of education's website.
- **Graduation Rate Incentive Award** – This new award program buttresses the state's focus on increasing graduation rates. The program is competitive and open to all Indiana public high schools. Monetary rewards go to the twelve public high schools with the greatest increases in graduation rates between the previous two school years: ten mid-to-large schools (more than three hundred students) receive \$20,000 awards, and two small schools (fewer than three hundred students) receive \$10,000 awards. The building principal and district superintendent of each winning school receive the funds and may distribute the cash awards to staff members whose work was critical to achieving the graduation-rate increases. The principal may receive no more than \$5,000 of the award amount.

Sanctions

Indiana's system for intervening in low-performing schools is relatively light on details and relies heavily on the resolve of the state and local boards of education. To begin, the state only holds schools accountable for performance; districts, no matter their performance designations, face no sanctions. The only schools that face state sanctions are those that repeatedly continue to earn F grades. For the first few years a school earns Fs, the system primarily requires local school boards to help create and implement improvement plans. The local school board can request that the state board of education appoint an outside team to manage the

school or assist in developing a new improvement plan, but it is not required to do so.

Not until years four, five, and six do the requirements become more prescriptive. At this stage, the state board appoints an expert team to rework the school's improvement plan and recommend changes. The expert team must include representatives from the community or region that the school serves. According to a representative of the state department of education, this requirement is intended to engage the community in holding schools accountable for meeting student learning needs.

If a school is still on academic probation in year six, the state board is required to conduct a hearing and consider testimony on options such as closing the school, merging the school with another school, or assigning a third-party Turnaround School Operator to run all or part of the school. Other options, including continued revision of the improvement plan, are also on the table.

As such, Indiana's accountability system only addresses its lowest-performing schools, and allows them to languish in low-performance for years before taking any significant actions—and even then, it provides avenues through which schools can avoid those drastic interventions. Thus far, however, the state board of education has proven steadfast in its resolve and not opted for the easy way out (improvement-plan stasis). The 2011-12 school year marks the first year that schools could have earned F grades for six years running, and seven schools did so. Of those seven, the board chose the turnaround option for five schools and a slightly less intensive option for the remaining two schools. Of course, it remains to be seen whether the turnarounds will work or if the board will opt for closure or school merges—presumably even more severe options—in the future.⁵⁸

Supports

As its relatively sparse outline of sanctions demonstrates, the Hoosier State's accountability system aims to drive improvement through progress monitoring and support, not through mandated interventions. On that side of the coin, Indiana provides its schools with a full kit of diagnostic and assessment tools to support local improvement efforts. See *Diagnostic and Assessment Tools* for a partial list of these tools made available by the state (some are required and some are optional).

One superintendent described how such tools support local school improvement efforts: “With the data from [diagnostic assessments], we’ve seen our buildings

⁵⁸ A state representative noted that the state is working to expand the outline of sanctions for schools earning both Ds and Fs, instead of just Fs; to eliminate the weaker options allowed for schools in year six; and to establish an outline for district accountability.

Diagnostic and Assessment Tools

IREAD-3 - The purpose of the Indiana Reading Evaluation And Determination (IREAD-3) assessment is to measure foundational reading standards. Based on the Indiana Academic Standards, IREAD-3 is a summative assessment that evaluates reading skills of third-grade students to ensure that all students can read proficiently before moving on to fourth grade. IREAD-3 is administered in addition to ISTEP+ and provides timely information in order to plan for summer intervention if needed, as well as for decisions relative to fourth-grade placements. IREAD-3 is mandatory.

mCLASS - The purpose of the mCLASS assessments is to provide diagnostic measures for K-2 students in literacy and numeracy. These assessments help identify students' foundational skills and provide teachers with instructional suggestions based on student performance on benchmark assessments and regular progress monitoring. These assessments are optional.

Acuity (3-8 and Algebra) - The purpose of the Acuity assessments is to provide diagnostic measures for students in grades three through eight in reading, mathematics, science, and social studies, as well as for students in Algebra I. Assessment reports provide standards-aligned performance data, which support educators' ability to inform instruction. These assessments are optional.

make real gains...Because the state provides these resources it saves us expense at the local level... It's a big help."

5) A system of rewards and consequences to drive improvement at the individual student level

Indiana requires that high school students pass end-of-course assessments in English 10 and Algebra I in order to graduate. In addition, beginning in spring 2012, all third-grade students will be required to pass the IREAD-3 assessment in order to be promoted to fourth grade (a similar requirement exists in Florida).

Like all states, federal NCLB regulations require Indiana to allow students in low-performing schools to attend other, high-performing schools. According to a state representative, however, because Indiana's low-performing schools are primarily

clustered in Indianapolis and Gary, few high-performing options exist for students in failing schools; the state does not enforce or encourage this option in any meaningful way. On the bright side, Indiana recently introduced a voucher program for students statewide. While the program is an option for low-income students, and not specifically those in low-performing *schools*, the overlap of those groups means that the program will in practice allow many students in low-performing schools to attend other private schools.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

In its 2011 legislative session, Indiana passed a landmark law requiring local districts to develop annual teacher and principal evaluations. The evaluations must be “significantly” informed by student achievement and growth measures and must differentiate performance across four categories: highly effective, effective, improvement necessary, or ineffective.

Principals will continue to be employed under one-year contracts. For teachers, districts must now incorporate evaluation results into their compensation structures and employment decisions. New teachers will automatically fall into a “probationary” category, and can be dismissed after just one ineffective rating. They are also eligible for “established” status, akin to receiving tenure, after earning three effective or highly effective ratings in a five-year period. Established teachers cannot be dismissed unless they receive two consecutive ineffective ratings, or three ineffective or “improvement necessary” ratings over a five-year period.

Indiana also recently introduced a competitive Excellence in Performance Award to provide districts with grants for establishing bonus programs for effective and highly effective teachers. To apply, districts must outline their new evaluation metrics and describe the expected impact of the bonus program on teacher recruitment, development, and retention.

What are the strengths and limitations of Indiana’s accountability system?

Strengths

Strong academic standards. Indiana’s content standards are rated among the best in the nation. The state routinely uses national-level content experts to develop and review its academic standards. The state also annually tests in all four key content areas—reading, math, science, and social studies.

Robust accountability metric. By combining measures of proficiency, growth, and college and career readiness, Indiana captures a detailed picture of student, school, and district performance. It also disaggregates the lowest-performing quartile of students to ensure that schools and districts are held to account for those students' performance. In addition, the A-to-F grading scale for schools and districts is straightforward and easily understood by policymakers, parents, and community members.

A full toolkit of instruments to monitor student progress. The state's investment in diagnostic and progress monitoring tools to support the work of school improvement is key. Educators can access data that pinpoint individual-student learning needs, and building administrators have tools and resources to monitor progress toward success on summative assessments. The state has also embraced the use of technology; for example, parents have direct online access to their students' data, including longitudinal results.

Teacher evaluation system that informs salary and employment decisions. Indiana's new teacher (and principal) evaluation legislation revamps the state's traditional salary structure and requires evaluations to include student performance and growth indicators. Further, the state actively encourages districts to adopt model systems by offering competitive grants for teacher bonuses.

Limitations

Limited system of consequences. While it measures district performance, Indiana's accountability structure ties no supports, sanctions, or interventions to district results. The system only meaningfully addresses those schools designated as the lowest performing for multiple years. While the state has clearly invested in progress-monitoring tools to support improvement efforts, its accountability system could benefit from a wider range of interventions and greater urgency in implementing those strategies with low-performing schools.

Final Word

Indiana's focus on growth and progress monitoring appears to be its greatest asset. Still, the system rests on the assumption that simply giving districts the data and diagnostic tools to track and monitor progress can spur academic improvement. Perhaps. But the interventions for low performers must be robust and the implementation swift. In Indiana, the choice of state-level interventions is limited, and the powers that be could choose to implement the weakest available option. If

interventions are to be targeted to the most egregious cases, state leaders must be vigilant and bold in their prescriptions.

Information on Indiana's education-accountability system was primarily drawn from interviews with state representatives, district representatives, and local stakeholders, as well as from the Indiana Department of Education website at www.doe.in.gov. Additional information was drawn from the National Council on Teacher Quality's *2011 State Teacher Policy Yearbook*.

MASSACHUSETTS

Overview

In 1993, via a grand (and bipartisan) bargain in which lawmakers supported an increase in education funding in exchange for staunch accountability measures, the Massachusetts legislature passed its Education Reform Act and kicked off almost two decades of accountability provisions in its education system.⁵⁹ In addition to establishing an accountability system driven by district intervention and action, the law introduced curriculum standards, linked them to student assessments, and required all high school students to pass those assessments in order to graduate. From the beginning, Massachusetts made it clear that districts—not the state—were to be responsible for improving low-performing schools, but that students were to have a stake in their individual performance as well.

Today, Massachusetts's standards and assessments are the bedrocks of a system regarded as one of the best in the land. The state is well known for setting high bars for its exams and for ensuring that students are held to those lofty expectations. Massachusetts has seen great success as a result: Students in the Bay State routinely score at the top of the heap on the National Assessment of Educational Progress (NAEP), and also outperform many of their international peers—even in high-performing Asian countries—on assessments such as PISA.⁶⁰

Over the last decade, Massachusetts has continued to tweak its accountability system in order to meet shifting federal policy and broaden the system's purview. In 2010, Governor Deval Patrick signed the Act Relative to the Achievement Gap into law, considered to be the most significant education legislation in Massachusetts since the 1993 law. Better known as the Achievement Gap Act, this legislation shifts responsibility for the lowest-performing schools onto the shoulders of both the district and the state (instead of relegating it to districts alone). In doing so, the legislation aims to provide those entities with additional tools and support for turning around the lowest-performing schools.

In addition, Massachusetts strengthened educator accountability in the system through June 2011 legislation that overhauled educator evaluation requirements.

⁵⁹ As described by a *New York Times* editorial at the time, the bargain “[cast] a generous supply of crumbs for the pack of special-interest mice scurrying around.” Editorial, “New state law is a starting point,” *New York Times*, June 20, 1993. For more on the provisions of the Education Reform Act, see the Education Reform Progress Report published by the Massachusetts Department of Elementary and Secondary Education here: <http://www.doe.mass.edu/edreform/edreformreport/erprogrpt597-1.html>.

⁶⁰ See the National Assessment of Educational Progress at <http://nces.ed.gov/nationsreportcard/>; and Eric A. Hanushek, Paul E. Peterson, and Ludger Woessmann, “Teaching Math to the Talented,” *Education Next* 11, no. 1 (2011): 10-18, <http://educationnext.org/teaching-math-to-the-talented/>.

Evaluations must now include objective measures of student performance. The state did not, however, tie evaluations to employment decisions or compensation, leaving teacher accountability in the state somewhat shallow.

Below, we map Massachusetts's progress against six key components of strong state accountability systems.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

Massachusetts boasts some of the nation's best standards—what the state calls “curriculum frameworks.” The state was celebrated for its reading and mathematics standards long before the Common Core in 2010—which is, in fact, why the decision to adopt the Common Core proved so controversial in the state. Massachusetts's U.S. History and science standards are also exemplary.⁶¹

In addition, Massachusetts prides itself on its high cut scores on the Massachusetts Comprehensive Assessment System (MCAS) tests, which are required for reading and math in third through eighth grades and again in tenth grade; for writing in grades four, seven, and ten; and for science and technology/engineering in grades five, eight, and ten.⁶² (Due to budgetary constraints, Massachusetts suspended its history and social science tests in 2009.) To receive a high school diploma, students must “pass the MCAS.” This means they must achieve proficiency on the state's tenth-grade reading and math MCAS tests, as well as meet a lower standard of performance on a subject-specific high school science assessment. (See page 72 for a more detailed discussion of the state's high school graduation requirement.)

⁶¹ In reviews by the Thomas B. Fordham Institute, Massachusetts received an A-minus for its U.S. History standards in 2011 and an A-minus for its science standards in 2012. The Common Core standards for reading and math, which Massachusetts adopted in 2010, earn grades of B-plus and A-minus, respectively. Previously, Massachusetts's state reading and math standards had earned grades of A-minus and B-plus, respectively. See *The State of State U.S. History Standards 2011*, *The State of State Science Standards 2012*, and *The State of State Standards—and the Common Core—in 2010*, at www.standards.educationgadfly.net/.

⁶² When states' proficiency cut scores are mapped onto the National Assessment of Educational Progress (NAEP), Massachusetts's cut scores are found to be the most rigorous in fourth-grade reading, fourth-grade math, and eighth-grade reading. (Massachusetts finds itself near the middle of the pack in eighth-grade reading.) Massachusetts is the only state to exceed the rigor of the NAEP cut scores in any grade or subject, doing so in both fourth- and eighth-grade math. See National Center for Education Statistics, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>. As Massachusetts has adopted the Common Core standards and currently sits on the governing board of the Partnership for the Assessment of Readiness for College and Careers (PARCC) assessment consortium, it will likely replace its MCAS tests with CCSS assessments in 2014–2015.

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Massachusetts provides a wealth of publicly available data on its department of education website. Through its school and district profiles database, users can access grade-level MCAS results, four- and five-year graduation rates, AP performance and participation rates, SAT results, mobility rates, and more, all disaggregated by subgroups including ethnicity, special education, English language learners, and low-income status. Further, the state reports the percentage of students at each of four proficiency levels for each school and district: Advanced, Proficient, Needs Improvement, and Warning/Failing. (“Warning” applies to students in grades three through eight, and “Failing” applies to high school students.) It also calculates Student Growth Percentiles (SGPs) to complement the MCAS proficiency scores. The SGPs measure individual student achievement over time rather than grade-level achievement results in a given year. SGPs are aggregated and reported by school and district as well. In addition to data on students, Massachusetts presents teacher information on licensing status, age, gender, race, and program area (general education, special education, English language learner, etc.).

Overall, while the database does not allow users to make easy school-to-school or district-to-district comparisons, for most indicators it does present individual school data next to results for that school’s district and the entire state. For those who are willing to dig, the state does provide a separate tool with which users can make school and district comparisons. The state provides the District Analysis and Review Tool (DART) to help districts, schools, and stakeholders meet accountability requirements. It allows users to compare schools and districts across a number of different indicators. The tool is publicly available—so any parent or community member may access it—but it is not prominently presented on the homepage of the website, as the school and district profiles are. Users must venture into the “accountability, partnership, and assistance” section of the website to find it; we may assume, then, that many don’t.

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

Massachusetts analyzes school, district, and subgroup MCAS data according to its Composite Performance Index (CPI). The CPI is a 100-point index that attaches a score to the percentage of students achieving Proficient in a given year. Consecutive CPI scores serve as the basis for determining a school or district’s AYP status, along with attendance, graduation rates, and test participation rates. This structure allows for relatively seamless integration of the state and federal accountability systems—a rarity across the states.

Based on CPI, AYP proficiency targets are set for each district, school, and student group indicating the amount of gain needed to stay on track toward 100 percent proficiency by the 2013-14 school year. Schools and districts can make AYP by meeting either the state's ELA and math proficiency targets (which were CPI scores of 95.1 and 92.2 in 2011, respectively) or their individual-group improvement targets. Individual targets are calculated by subtracting the district, school, or subgroup's current CPI from 100 (the ultimate CPI goal), and dividing by the total number of years remaining until 2013-14.⁶³ The establishment of differentiated gain targets sets Massachusetts apart from the many states that set uniform targets for all of their schools and districts. Superintendents report a sense of "fairness" at the local level since improvement targets are based on the prior performance of each school, rather than a "one size fits all" approach. Still, one superintendent noted that while the individual targets have traditionally allowed urban schools that fall short of state benchmarks but make progress to make AYP, an increasing number are now missing these individual benchmarks as well.

In addition, Massachusetts uses these data to assign schools and districts "Improvement Ratings" based on a comparison of consecutive years' CPI results. Ratings indicate whether a school or district is Above Target, On Target, Improved, Below Target, Made No Change, or Declined. While inherently tied to AYP, these ratings are descriptive only (i.e., they alone require no action toward improvement).

As described in more detail below, schools and districts are also assigned to School and District Accountability and Assistance Levels—from Level 1 (highest-performing) to Level 5 (lowest-performing)—that entail state-required improvement actions beyond those required by NCLB. These designations are based on consideration of both objective measures of AYP/test scores and subjective state reviews.

Ultimately, then, schools and districts are scored against three different rating systems: AYP status, Improvement Ratings, and School and District Accountability and Assistance Levels. While Massachusetts's integration of state and federal accountability eliminated much redundancy and confusion inherent in other states' systems, superintendents report that these overlapping labels are confusing to parents.

⁶³ District-level CPI calculations are issued separately for elementary, middle, and high school grade spans. Under this approach, districts are only identified for improvement when they fail to make AYP in the same subject area in all grade spans.

4) A system of rewards and consequences to drive improvement at the school and district levels

Rewards

Budget cuts have recently led Massachusetts to eliminate programs that monetarily reward high-performing and/or improving schools. Previously, the state recognized schools demonstrating significant improvement in reading and mathematics by designating them as Commonwealth Compass Schools and providing them with financial awards. Through a similar program, Sustaining Success Grants, the state financially rewarded schools that made AYP for two consecutive years. These funds were intended to support the continuation of strong instructional practices and initiatives in individual schools.

At this time, no incentive programs exist to recognize or promote school improvement other than additional Commendation Designations on the school's accountability report. These note that a school is "commended for" narrowing proficiency gaps, achieving high growth, and/or exiting NCLB accountability status. As one mid-size district superintendent said, "Our high school [achieved] substantial success and all we got was an 'atta-boy.' The system is all stick and no carrot."

Sanctions

To determine required improvement actions, Massachusetts assigns each school and district to an Accountability and Assistance Level, with the highest-performing in Level 1 and the lowest-performing in Level 5. For both schools and districts, Levels 1-3 are based on objective measures of student performance; Levels 4 and 5 also include measures of student performance, but schools are ultimately assigned to them based on state discretion.

It is important to note that undergirding the school and district Accountability and Assistance Levels is the state's *Framework for District Accountability and Assistance*. The framework requires particular actions on the part of the district and the state depending on the Accountability and Assistance Level to which a *district* is assigned. So while *schools* are also assigned to Accountability and Assistance Levels, those are primarily used to identify low-performing schools in order to determine district Accountability and Assistance Levels. Thus, school designations are not tied to school-driven accountability actions. Instead, the state places districts in the driver's seat of accountability, with the state primarily building district capacity and only playing a more direct role in the improvement of the state's very worst districts.

School Accountability and Assistance Levels are assigned as follows:

- Level 1: Schools making AYP or in their first or second year of improvement.
- Level 2: Schools in NCLB corrective action or restructuring.
- Level 3: The lowest-performing 20 percent of all schools statewide, based on the percentage of students scoring at the Warning/Failing level on MCAS tests, CPI, median Student Growth Percentile (SGP), and for high schools, dropout and graduation rate.
- Level 4: Schools among the lowest-performing 4 percent of all schools statewide (based on the measures above) that require intensive intervention, **as determined by the state**; these are otherwise known as “turnaround schools.”
- Level 5: Those turnaround schools for which Level 4 status did not yield sufficient improvement, **as determined by the state**.

District Accountability and Assistance Levels are similarly defined. Level 1 includes those districts with no schools or subgroups in NCLB corrective action or restructuring, while those with at least one school or subgroup in corrective action or restructuring are placed into Level 2. In both of these Levels, districts must review and revise their own improvement plans, while the state may only provide voluntary improvement tools (Level 1) and suggest improvement strategies (Level 2). Level 3 includes those districts with at least one school among the lowest-performing 20 percent of schools in the state (i.e., at least one school in Level 3). Here, districts must use the state’s self-assessment process to review and revise their improvement plans, while the state must offer priority—but again voluntary—assistance to districts.

At each of these District Accountability and Assistance Levels, the state conducts “random” (Levels 1 and 2) or “selective” (Level 3) district reviews. Based on the results of these audits, the state can choose to place a district into Level 4, if it determines that at least one of the district’s schools should be a Level 4 turnaround school, or if it determines the district as a whole requires state intervention. Here, the district must work with the state (as well as teachers, administrators, local stakeholders, parents, and teacher unions) to develop an improvement plan for the district as a whole, including turnaround plans for any schools in Level 4.

If the district fails to meet progress benchmarks outlined in the improvement plan, or if the state determines that the district requires further intervention, the state may place the district into Level 5. Here, the state requires that a third party, or “receiver,” appointed by the state share “co-governance” of schools with the district and jointly determine major budgetary, personnel, and policy decisions. The receiver, however, may be granted powers up to and including those of the superintendent, and thus in some cases may exercise full governance of the

district. No more than three districts can be designated Level 5 at any one time; for 2011-12, only one district was designated Level 5.

According to one stakeholder, the state review process ensures that schools and districts are identified for intervention (i.e., placed in Levels 4 and 5) based on more than just student test scores, as the reviews also examine finance, governance, professional development, and so on. But at the same time, the subjectivity of the process allows the state to identify only as many Level 4 schools as it has the capacity to assist—and it caps that total at the worst-performing 4 percent of schools. As a result, many low-performing schools that should be identified as Level 4 wind up in Level 3, and many other low-performing, but slightly better, schools go unnoticed.

Supports

Massachusetts's *Framework for District Accountability and Assistance* places primary responsibility for improvement efforts on the shoulders of districts; that said, the state provides direct and indirect support to districts, increasing in substance through each of the five Levels.⁶⁴ For Levels 1-3, the state offers voluntary tools and resources for districts to use in self-assessment, professional development, and improvement planning. For example, the online District Analysis and Review Tool allows districts, schools, and stakeholders to make quick district and school comparisons across indicators of student performance, graduation and dropout rates, discipline violations, leadership turnover, finances, and more.

In addition, districts in Levels 3 and 4 are given priority access to six regional District and School Assistance Centers (DSACs) that aid districts in self-assessment, data analysis, and improvement planning, as well as in identifying/sustaining effective practices and utilizing state-provided tools. (Districts in Levels 1 and 2 may also work with these regional networks, to the extent that enough resources are available.) District superintendents highlighted the value of the assistance centers at the local level and found their support to be very helpful, noting that these organizations are clear that their primary role is to assist schools and districts and not to *direct* improvement efforts.

Also for districts in Level 4, the state appoints an “accountability liaison” to each district to coordinate implementation of the intervention actions outlined above. In addition, these districts can apply for School Redesign Grants for each of their

⁶⁴ While not technically components of the state's accountability system, nonprofits such as Mass Insight and the Massachusetts Business Alliance for Education have also provided much support to districts and schools over the last few decades, offering field services and promoting standards-based reform and leadership development. Nonprofits have also played a role in helping to develop many of the state's current accountability provisions.

Level 4 schools; these grants are competitive, however, and are distributed based on the quality of the schools' turnaround plans. State support for schools in Level 5 comes in the form of the assigned "receiver."

Finally, the state has identified ten "Commissioner's Districts," which are generally large, urban centers with high poverty levels and low-performing students. The state provides additional resources and tools to these districts to target and improve their lowest-performing schools and support their remaining schools.

5) A system of rewards and consequences to drive improvement at the individual student level

Through its well-regarded state assessment system, Massachusetts aims to hold its students accountable for their own performance. That said, specific provisions aimed at students are somewhat limited in scope, and Massachusetts offers no rewards for students who perform at high levels. Further, by establishing "safety nets" to ensure that every student can demonstrate proficiency in some way, the Bay State potentially weakens what would otherwise be strong student-level accountability.

To receive a high school diploma, students must ostensibly pass the MCAS by meeting or exceeding the Proficient level (a score of 240) on the tenth-grade English language arts (ELA) and math MCAS tests, as well as meeting or exceeding the Needs Improvement level (a score of 220) on a subject-specific MCAS science test (Biology, Chemistry, Introductory Physics, or Technology/Engineering).⁶⁵ But in reality, as one stakeholder put it, these requirements are "aspirational": In 2006, the Board of Education voted to raise the MCAS cut scores for ELA and math from 220 to 240, as many students were meeting the 220 thresholds in tenth grade and then not continuing on in their coursework (particularly in math); thus the decision was an attempt to push more students to take additional courses in core subject areas even after they'd met the state requirement. But, along with that decision, the board introduced an alternate path to graduation—the Educational Proficiency Plan (EPP). If a student meets or surpasses a score of 220 in both ELA and math, but fails to achieve the 240 cut score in those subjects, he is placed on an individual EPP, which is developed by the district and specifies the courses that he will be required to pass in eleventh and twelfth grades, as well as the assessments the school will use to monitor his progress toward proficiency. These assessments include the ELA MCAS retest, the mathematics MCAS/EPP test, or a

⁶⁵ The state plans to fold history and social science assessments into the graduation requirement once those tests—which were suspended in 2009, due to budgetary constraints—are reinstated in the future.

locally designed assessment.⁶⁶ Thus, in reality, students on EPPs who have failed to reach 240 on the MCAS once are technically not required to take it again, and can graduate without ever meeting that specific provision. While the EPP process outlines other requirements that students must meet, one superintendent noted that the EPP process is “tough to do with fidelity without time and money.” Without proper attention paid to ensuring that it upholds a high bar, the EPP process *could* be an easy graduation loophole for low-performing students—and one stakeholder concurred that many students who never achieve the rigor of the 240 cut score continue to receive diplomas.

In terms of supports, Massachusetts does offer students stuck in low-performing schools a handful of opportunities to move to other schools. But the state does not adequately encourage students to do so. While it allows districts to participate in inter-district school choice—and while most districts in the state participated to some extent in 2011, either sending or receiving students—only about 1 percent of all students opted to transfer to schools out of their home districts in 2011-12. The state also allows districts to offer intra-district choice, but the extent to which districts choose to do so—and the extent to which students take advantage of this option—is questionable.⁶⁷ Finally, the state established the METCO program in 1966, allowing students in Boston (and later, those in Springfield) to transfer to nearby suburban districts. But this program is very small, with only about 3,000 participating students each year, despite its long waiting list. So while Massachusetts does have some transfer options on paper, one stakeholder bluntly replied that in practice, “no notion of student choice options” exists statewide.

Massachusetts does, however, have a growing charter school student population, particularly in its urban areas. More than 30,000 students were enrolled in charters at the outset of 2011-12 (more than double the number attending out-of-district schools), and over 35,000 more were on charter waiting lists.

⁶⁶ Eleventh- and twelfth-grade students who have not yet scored 240 on the ELA MCAS test may take the ELA MCAS retest. For mathematics, only students who score below 220 may retake the MCAS in that subject. Those scoring above 220 must take the mathematics MCAS/EPP Test, which is aligned to the state’s content standards but, unlike the mathematics MCAS, includes only multiple choice questions (no short answer or constructed response questions).

⁶⁷ The most recent data to our knowledge were published in 2003 by the Center for Education, Research, and Policy at MassINC. Analysts report that twenty-one districts provided intra-district school choice in 2003; however, many districts are constrained by the small number of schools within the district. Of the 184 districts serving middle school students, 142 had just one middle school; and of the 225 districts serving high school students, 206 had just one high school. See *Mapping School Choice in Massachusetts: Data and Findings 2003* (Boston, MA: MassINC, May 2003), http://www.renniecenter.org/research_docs/o305_SchoolChoice_report.pdf.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

As promised in the state's 2010 Race to the Top application, Massachusetts adopted new educator evaluation requirements—applying to teachers, principals, and superintendents—in June 2011. Evaluation systems are still collectively bargained at the district level, but Massachusetts has now established broad criteria that evaluation systems must meet, and the state offers a model evaluation system which districts can choose to adopt or adapt. Unfortunately, while the new provisions mark a step forward for adult accountability in the Bay State, the requirements don't go far enough toward linking educator and student performance and tying evaluations to employment decisions.⁶⁸

Evaluation systems now comprise two separate ratings—an overall rating and an “impact on student learning” rating. Thus, while emphasis is placed on measures of student learning, those measures are not required to be the primary criteria of the overall evaluation.⁶⁹ Overall ratings differentiate performance across the following four categories: Exemplary, Proficient, Needs Improvement, and Unsatisfactory. Student-learning ratings designate whether an educator has had high, moderate, or low impact on student learning. Based on their overall and student-learning-impact ratings, educators are placed on different types of “Educator Plans” that are intended to provide feedback and specify actions to be taken toward specific goals. Any educator rated as Unsatisfactory must be placed on an “Improvement Plan.”

Teachers are generally evaluated annually, with the exception of tenured teachers with Exemplary or Proficient ratings who also have moderate or high impacts on student learning; they may be evaluated every two years. Teachers are eligible for tenure after three years. To earn tenure, a teacher must receive ratings of Exemplary or Proficient on all evaluation categories—including student learning—as well as on the evaluation as a whole. A principal may confer with the superintendent if he or she wishes to award tenure to a teacher who does not meet these criteria.

Massachusetts does not tie evaluations to compensation. The state sets a minimum teacher salary, but allows districts to design their own compensation

⁶⁸ Beyond evaluation and employment decisions, one stakeholder in Massachusetts said a key lever of education reform is the state's requirement that teachers pass rigorous licensing tests aligned with the state's content standards. But according to the National Council on Teacher Quality, the state leaves room for improvement in terms of ensuring its educators are well prepared because the licensing tests do not always score different subjects separately. Thus, a teacher could mask poor performance in one subject with strong performance in another. See the *2011 State Teacher Policy Yearbook—Massachusetts* (Washington, D.C.: National Council on Teacher Quality, January 2012), http://www.nctq.org/stpy11/reports/stpy11-massachusetts_report.pdf.

⁶⁹ Other measures of teacher performance must include administrator and student feedback, in addition to materials provided by the teacher. The state is currently examining the feasibility of including parent feedback in teacher evaluations as well.

systems. Further, the state only loosely ties evaluations to dismissal, by specifying that educators who fail to meet the conditions outlined in their Improvement Plans *may* be eligible for dismissal—but not requiring that persistent low-performers be dismissed. In addition, tenured teachers who are dismissed may appeal multiple times. Layoffs are also not linked to teacher performance—the state stipulates that districts may not lay off a tenured teacher if a probationary teacher within the same certification area could be laid off instead. In short, the state does not prohibit “last in, first out” policies—rather, it seems to encourage them.

Beyond consequences for low-performing educators, Massachusetts offers only limited rewards for high performers. The state’s Improving Educator Quality grants provide funds to districts for additional pay for teachers who work in high-needs schools or certain subjects. But these funds are limited, and as mentioned above, the state does not require salary to be based on performance.

Finally, there’s some attention to accountability for administrators—but not much. Building leaders in their first three years in a district must be evaluated annually. Administrators are hired on one-year contracts.

What are the strengths and limitations of Massachusetts’s accountability system?

Strengths

Strong academic standards and rigorous cut scores. Massachusetts’s academic standards are rated among the nation’s best in all core subject areas. In addition, the state has not shied away from setting high cut scores on its assessments.

Integration of state and federal accountability systems. Massachusetts has integrated its state accountability system with federal NCLB requirements, and as a result has minimized confusion at the district and school level. Absent from district interviews were the negative comments and confusion heard from multiple insiders in other states that have bifurcated systems.

Differentiated progress targets. Massachusetts sets differentiated improvement targets for each school, district, and student group indicating the amount of gain needed to stay on track to 100 percent proficiency by the 2013-14 school year. Schools have the option of either meeting these differentiated improvement targets or the state proficiency target. District administrators report that, as opposed to the state proficiency target, the differentiated targets “feel relevant and achievable.”

A district-driven framework for accountability and assistance.

Massachusetts's Accountability and Assistance Levels delineate district and state responsibilities for high- and low-performing schools and districts. To the extent that the two entities share responsibility for low-performing schools, the framework makes clear that the district should remain, to the greatest extent possible, the driver of school improvement.

Accountability provisions for students. By setting high cut scores on the state assessment, and by requiring all students to pass the tenth-grade test in order to graduate, Massachusetts ensures that students have a stake in their own education. But as colleges in the state do not accept the graduation requirement as a measure of admissibility, Massachusetts would do well to ensure that its high school assessments are aligned with measures of college and career readiness.

Limitations

Limited indicators to inform performance designations. The indicators that Massachusetts employs to designate school and district performance are limited to measures of proficiency and attendance, graduation, and test-participation rates. While Massachusetts calculates and reports Student Growth Percentiles (SGPs) to measure individual student achievement, these data only factor into the accountability system when the state is determining whether to rank a school or district in Accountability and Assistance Level 3, 4, or 5.

No concrete rewards for high-performing schools and districts. Recent budget reductions have eliminated all performance-based incentives except labels that designate a school or district as high-performing in a particular area.

Limited number of schools identified for intensive interventions. As Massachusetts only identifies for intervention as many schools for which it has the capacity to intervene, the percentage of all schools that are required to undertake robust improvement actions is extremely small—below 4 percent. Districts are largely left to attend to other struggling schools. While we can appreciate the reality of limited dollars, a system which lays out a more structured, increasingly extensive series of interventions—supported by the state, district, nonprofits, or other groups—might better address schools along the low-performing continuum.

Few policies to drive teacher and administrator performance. While Massachusetts took a big step forward in overhauling its educator evaluation requirements and ensuring that student performance factors

into them, the state does not require that such evaluations be used to inform compensation or employment decisions for either low or high performers. The Bay State should take a page out of its student policy playbook and ensure that teachers have a clear stake in their own performance—and in the performance of their students.

Final Word

Massachusetts's accountability system keeps much responsibility for performance at the local level—primarily holding districts and students accountable for low performance—but the state has done its share to ensure that both of those entities are held to high standards. A more robust structure for addressing low-performing—but not lowest-performing—schools, coupled with strengthened educator accountability measures, would go a long way toward making accountability in the Bay State a model for the nation.

Information on Massachusetts's education-accountability system was primarily drawn from interviews with state representatives, district representatives, and local stakeholders, as well as from the Massachusetts Department of Elementary and Secondary Education's website at <http://www.doe.mass.edu/>, particularly the website's Accountability/Assistance page at <http://www.doe.mass.edu/Assess/>. Additional information was drawn from the National Council on Teacher Quality's *2011 State Teacher Policy Yearbook*.

OHIO

Overview

In the summer of 1997 the Ohio General Assembly passed Senate Bill 55, introducing Ohio's first state-level accountability system. Since then, the Buckeye State's system has continued to evolve, often in response to the push and pull of politics and funding. In 2011, for example, during a landmark legislative session that saw Governor Kasich try—and fail—to overhaul collective-bargaining provisions, the legislature took steps to strengthen Ohio's accountability measures. House Bill 153, the biennial budget bill, introduced stricter turnaround and closure requirements for schools, a pilot “parent trigger” option for parents with children in low-performing schools, and new teacher-evaluation requirements, among other elements.

Senior-level state education staff indicate that Ohio's existing accountability system has three primary objectives: (1) establishing and communicating to districts clear expectations about what students should know and be able to do; (2) informing the public about school performance through transparent school-level achievement data; and (3) providing school administrators and teachers with an objective set of data to drive decision-making and improvement planning. Each of these objectives has been met with varying levels of success, as explained below.

Below, we map Ohio's progress against six key components of strong state accountability systems.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

The quality of Ohio's reading and math standards markedly improved with the state's adoption of the Common Core in 2010. The Buckeye State would do well to refurbish its U.S. History standards in a similar manner and to tweak its science standards, which, though respectable, leave some room for improvement.⁷⁰ The Ohio Achievement Assessments annually assess reading and mathematics in third through eighth grade and again in tenth grade; science in fifth, eighth, and tenth grades; and writing and social studies in tenth grade. (In 2009-10, Ohio discontinued its fifth- and eighth-grade social studies tests and its fourth- and seventh-grade writing tests due to financial constraints.) The tenth-grade tests together comprise the overall Ohio Graduation Test (OGT).

⁷⁰ In a 2010 analysis by the Thomas B. Fordham Institute, Ohio's reading and math standards both received Cs, while the Common Core reading and math standards received a B-plus and an A-minus, respectively. The state received a D for its U.S. History standards in 2011, and a B for its science standards in 2012. See *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, and *The State of State Science Standards 2012*, at www.standards.educationgadfly.net/.

Unfortunately, Ohio does not set a high bar for its assessments. When mapped onto proficiency scales employed by the National Assessment of Educational Progress (NAEP), Ohio's standard of proficiency on its fourth- and eighth-grade math and eighth-grade reading assessments only matches NAEP's *basic* level; its fourth-grade reading cut score only meets NAEP's *below basic* level.⁷¹ As one district representative reported, this problem persists at the high school level: "There is a challenge [in] trying to move kids to a more demanding curriculum [when] they are assessed at a lower level with the OGT."

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Ohio reports school and district performance using a web-based interactive Local Report Card (iLRC). The database provides a level of transparency not typically found in most state reporting systems, and its interface is simple and intuitive. The system allows for quick comparisons of schools and districts on a number of indicators—such as test results, graduation rates, student discipline, mobility rates, grade-level promotions, and school and district ratings under the state accountability system, among other elements. Much of the data can be disaggregated by student subgroups—such as race, disability, and homeless status. Further, the proficiency data are presented by the percentages of students scoring at different proficiency levels, rather than by a single rate of proficiency.

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

To assign performance ratings to its schools and districts, Ohio measures schools based on both state and federal indicators and produces school and district report cards which assign each entity one of six designations:

- Excellent with Distinction
- Excellent
- Effective
- Continuous Improvement
- Academic Watch
- Academic Emergency

⁷¹ National Center for Education Statistics, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>. Ohio is a governing state of the PAARC assessment consortium and will likely replace its reading and math assessments with the Partnership for Assessment of Readiness for College and Careers (PAARC) assessments in 2014-15.

Performance designations for schools and districts are based on the combination of the following four distinct measures:

State Indicators – The State Indicators report measures the percentage of all students scoring proficient or above on the Ohio Achievement Assessments (OAA) and the Ohio Graduation Test (OGT). The passing standard for schools requires that 75 percent of their students pass each content area test, and that 85 percent pass the OGT by the end of eleventh grade. The State Indicators report also measures whether schools are meeting the required attendance and graduation rates, which are 93 and 90 percent of all students, respectively. (These data are not disaggregated by subgroup.)

Performance Index Score (PI) – Districts and schools earn points based on how well each student performs on the OAA and the OGT. The performance index score is a weighted average that includes student test results in all tested subjects and grades. The greatest weight is given to advanced scores; weighting decreases for each performance level and a weight of zero is given to untested students. This results in a scale from 0-120 points. The PI can be compared across years to show school achievement trends. Districts report the PI as a particularly valuable metric as it gives them a “big picture look of the schools’ overall health over time.”

Value-Added Measure – In 2007, Ohio added a value-added measure to capture student progress at the school level from one year to the next. Value-added results may be used to augment an overall performance designation. Since they demonstrate growth rather than absolute performance, their inclusion functions as a reward for low-performing schools that are making substantial progress but not hitting absolute performance standards. The value-added measure has great promise, but Ohio needs to build district confidence in the measure’s accuracy (especially if it is going to serve as a foundation for proposed pay-for-performance compensation systems). More than one administrator expressed concern about its accuracy and felt that in-district benchmark assessments were more reliable gauges of student progress and teacher effectiveness.⁷²

Adequate Yearly Progress (AYP) – Every school and district in Ohio must meet AYP benchmarks for reading and math proficiency as well as for test participation, attendance, and graduation rates. Ohio calculates proficiency rates of student groups by tallying both the number of students proficient

⁷² One state representative, however, countered that Ohio has already acted to strengthen the inferences that can be drawn from its value-added results. The state has implemented a smoothing process to address the annual “expected” fluctuations on the assessments and changed the standard error of measurement.

in a subject *as well as the number of students on track to reach proficiency in that subject within two years*, and then dividing by the total number of students.⁷³ As one local stakeholder put it, this measure inflates the proficiency rates of Ohio's student population, and is often considered to be one of the worst AYP proficiency measures in the nation.

Assigning school and district designations is a two-step process. First, State Indicators, Performance Index scores, and AYP are combined to form a preliminary designation. If consideration of AYP bumped a school or district up or down a designation level, then that preliminary designation serves as the final designation. If not, the state then incorporates value-added data into the calculation, which may increase (but not decrease) a school or district's final designation.

So while Ohio's performance designations incorporate multiple measures of achievement, the structure of the calculations can result in an inflated picture of student performance. For one, the state's AYP proficiency measure is broad to the point of meaninglessness. Compounding that with value-added scores that can be used to bump up (but not down) schools and districts makes for a potentially inaccurate identification system that may mask areas of weakness.

Further, Ohio's performance labels are just that—labels. As described below, Ohio only ties these designations to sanctions and supports for the very worst-performing schools. Otherwise, Ohio relies on AYP status and NCLB requirements to hold schools accountable.

4) A system of rewards and consequences to drive improvement at the school and district levels

Rewards

According to local stakeholders, the performance rating system is an important “carrot and stick” for schools in Ohio. When asked about the importance of incentives, one superintendent responded, “The biggest incentive is the community expectation to pursue or keep the rating of Excellent or Excellent with

⁷³ The proficiency standards for AYP are different from the standards for determining proficiency in the State Indicators, PI, and value-added calculations. In addition, AYP disaggregates data by student subgroups, while the State Indicators, PI, and value-added measure do not. Thus, while the combination of these indicators produces a robust picture of school and district performance, the different state and federal measures can contribute to general confusion about the overall metric. For example, the accountability system requires that a district or school that meets AYP can be designated no lower than Continuous Improvement, while an Excellent or Effective district or school that does not meet AYP goals for three or more consecutive years (and in the same two or more subgroups) will see its designation reduced by one category. This can lead to final designations that do not appear to align with the state-required indicators. At least one Ohio superintendent wishes that the state would “report [AYP] but not include it as a component of the campus/district rating.”

Distinction.” Not only does Ohio provide the “stick” for schools that don’t achieve these ratings (described below), it provides the “carrot” for those that do—an additional \$17 more per pupil in FY 2012-13 for those districts ranked as Excellent or Excellent with Distinction. In addition, while the provision of additional funds based on performance is a relatively new development, districts with those designations have long had the ability to seek waivers to opt out of state requirements.

Above and beyond the six-level designation system, Ohio reports a list of buildings deemed the State Superintendent’s Schools of Promise—an award given to economically disadvantaged schools that have met standards for AYP, graduation rates, value-added measures, and student proficiency. This designation, however, includes no financial reward.

Sanctions

Ohio has traditionally maintained few school and district sanctions beyond those mandated by NCLB, and even then the state permits its schools some leeway in adhering to them. Since 2008, Ohio has participated in a federal pilot program that allows it to implement a *Differentiated Accountability* model; it permits the Buckeye State more flexibility in meeting federal accountability requirements. In practice, schools and districts can avoid implementing many federal improvement efforts entirely, and instead craft their own improvement plans under state supervision.

Here’s how the *Differentiated Accountability* model works: Ohio requires that all schools and districts in NCLB “school improvement” (meaning those that have missed AYP for at least two consecutive years) complete the Ohio Improvement Process (OIP). The process requires state teams and Educational Service Centers (ESCs)⁷⁴ to work with districts and schools to identify areas for improvement, develop improvement plans, and implement and monitor the plans. Throughout the process, the state provides districts with state-developed products and tools, including professional development. The state also provides training in the OIP process for ESC personnel and others who function as regional facilitators, ensuring that districts receive consistent support in their improvement efforts.

Potentially more important to districts than these supports, however, OIP also removes the requirement that districts spend their time in sweeping corrective action—such as replacing building staff and restructuring the internal organizational structures—which some leaders see as unrelated to their daily headaches. The OIP has received national recognition in that regard, as it

⁷⁴ Educational service centers (ESCs) work under the direction of and in coordination with the state education agency. According to the ODE website, ESCs provide services to assist districts and schools in improving student achievement and operating more economically and efficiently.

addresses the complaint that NCLB requirements are indiscriminate and disjointed from schools' and districts' actual needs, and at least one school superintendent indicated that the process showed promise in its early stages. But in practice, the OIP has become an institutionalized method for avoiding NCLB requirements. Superintendents described it as a “compliance activity” and a “box-checking activity,” that simply allows NCLB sanctions to go away for schools that complete it.

However, Ohio's 2011 budget bill, H.B. 153, layered additional sanctions for Ohio's lowest-performing schools onto these flimsy “differentiated” requirements. The legislation mandates the restructuring or closure of schools that are in the lowest 5 percent of all schools based on PI scores for three consecutive years and that are rated Academic Emergency or Academic Watch. The four turnaround options outlined by the state are stricter than those outlined federally under the School Improvement Grants (SIG)—for example, Ohio's turnaround options do not include the SIG “transformation” model, which typically leaves the majority of a school's staff and programs in place.

In addition, the budget bill included a “parent trigger” provision to be piloted in Columbus City Schools and potentially expanded statewide. The bill targets parents with children in buildings that have ranked in the lowest-performing 5 percent of schools for three consecutive years. Those parents can petition a district to close and reopen the school as a charter school, replace at least 70 percent of the school's staff, contract with another school district or entity to operate the school, turn operation of the school over to the state, or make other fundamental reforms to staffing and governance (similar to NCLB's cascade of sanctions).

The addition of these two accountability measures is a big step forward for Ohio's otherwise weak accountability system. Still, there is some ambiguity as to how Ohio's new turnaround requirements will overlap and function in tandem with the federal system, particularly for those schools that have received SIGs under NCLB.

Support

Ohio's *Differentiated Accountability* model, described above, allows the state to offer different levels of support to schools and districts based on a more nuanced picture of low performance. Instead of focusing on the number of years that a school or district misses AYP—the traditional method for distributing support under NCLB—the model permits Ohio to base support on the aggregate percentage of a school or district's student groups that do not meet AYP in reading and mathematics. This is an improvement over the traditional NCLB support system, which requires that two schools that have missed AYP for an equal number of years—one missing AYP in all subgroups, and the other missing AYP in just one—receive the same level of support.

In addition, and like the OIP process itself, the state provides mechanisms through which districts and schools can obtain outside counsel and guidance. Ohio mandates that the superintendent of public instruction establish an Academic Distress commission when a district is designated to be in a state of Academic Emergency and has failed to make AYP for four or more consecutive years. The commission, which includes two community representatives appointed by the local school board president and three state appointed representatives, is charged with assisting the district in improving academic performance. At the time of this publication, only one district, Youngstown, was declared to be in Academic Distress.

In addition, Ohio requires that its lowest performers participate in the State Diagnostic Team Review Process. Designed for schools or districts designated high support (see below) with multiple curriculum areas, grades, and student groups not meeting AYP, this process requires skilled reviewers from outside the district or school to evaluate district processes, resulting in a diagnostic profile of needs-improvement areas for the district. The state describes the methods and protocols created for this review process as “grounded in scientifically based research practice, and aligned with the themes that have emerged from Ohio Schools of Promise case studies.”

5) A system of rewards and consequences to drive improvement at the individual student level

Ohio has some student-level accountability measures on the books, but these are not as stringent as they could be. For one, the state requires that students pass the Ohio Graduation Test (OGT) in order to graduate.⁷⁵ But because these tests are first administered in tenth grade, and can be retaken by students who do not pass, they are based on tenth-grade learning standards. As a result, district-level representatives express some concern that students are not held to a high standard for graduation. As one district representative opined, “The current indicators that measure proficiency and put students in proficiency categories are just a minimum.” Many district representatives look forward to the fundamental shifts in standards and assessments that will occur as the state implements the Common Core standards.

⁷⁵ A student may also graduate if he meets the following criteria: (1) he has passed four of the five tests and has missed passing the fifth test by no more than ten scale score points; (2) has had a 97 percent attendance rate each of the last four years and has not been expelled in the last four years; (3) has a grade point average of 2.5 out of 4.0 in the subject area missed and has completed the curriculum requirement in the subject area missed; (4) has participated in an intervention program offered by the school and has had a 97 percent attendance rate in any program offered outside the normal school day or year, including those offered by someone other than the school; and (5) has obtained letters of recommendation from each teacher in the subject area not yet passed and from the high school principal.

On paper, Ohio also requires that third-grade students pass the state's third-grade reading test in order to be promoted to fourth grade. In practice, however, Ohio does little to enforce this legislation; while about 20 percent of all third graders failed to achieve proficiency in 2009-10, only 0.6 percent of third graders were retained for 2010-11. The state is supposed to provide those students who failed but then advanced to fourth grade with additional reading supports, but there is little evidence that those additional supports have any positive impact.⁷⁶

Ohio does offer a number of options for its students to exit low-performing schools and choose schools that match their needs. Since 2006-07, the Educational Choice Scholarship Program has offered private-school vouchers to students in persistently failing schools (i.e., schools earning ratings of Academic Watch and/or Academic Emergency for two of the last three years). The state recently expanded this option to include students in schools ranked in the bottom 10 percent statewide for two of the last three years based on PI score.⁷⁷ Ohio also offers separate voucher programs for students with autism and for special-needs students writ large. In addition, Ohio allows for open enrollment among districts, expanding the options for students to exit low-performing schools. While not specifically limited to low performers, the policy allows districts either to permit students to enroll from any district in the state or to permit students to enroll from adjacent districts. About 80 percent of districts allow some form of open enrollment.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

Ohio included legislation in H.B. 153 introducing new requirements for teacher and principal evaluations. The state now requires that 50 percent of teacher and principal evaluations be based on student growth measures. Teachers and principals are to be rated according to the following categories: accomplished, proficient, developing, and ineffective. Districts must evaluate new teachers (i.e., non-tenured teachers) twice a year and all other teachers annually.⁷⁸ (The bill is agnostic on the frequency of principal evaluations.)

The legislation leaves many decisions to districts, however, meaning that the quality of implementation and the consequences tied to the evaluation system depend on district initiative. While the legislation requires that the evaluations

⁷⁶ Emmy L. Partin, "Ohio Needs a Reading Guarantee," *The Ohio Gadfly*, November 2, 2011, <http://www.edexcellence.net/publications-issues/gadfly/ohio/2011102-oh-gadfly.html>.

⁷⁷ In addition, proposed legislation to create the Parental Choice and Taxpayer Savings (PACT) Scholarship is currently working its way through the Ohio statehouse. The new voucher program would expand voucher eligibility to families with annual incomes less than \$95,000. The legislation passed the House education committee in September 2011 but must still gain full approval of the Ohio House and Senate. Local stakeholders noted that the legislation is not expected to pass, at least in its current form.

⁷⁸ With one exception: Districts can opt to evaluate "accomplished" teachers biennially.

“inform” key employment decisions, and that teacher layoffs can no longer be based primarily on seniority, the state does not stipulate whether the evaluations must be tied to dismissal, tenure, transfer, or compensation determinations for teachers or principals. (Only those districts participating in Race to the Top are required to develop teacher performance-pay scales.) Districts are required to have new evaluation systems in place by 2013.

Finally, H.B. 153 mandates that teachers of core subjects in schools ranked in the bottom 10 percent of schools each year must retake all of the state’s teacher licensure exams required for the grades and subjects they teach (not more than once every three years). While the measure seemingly aims to hold teachers accountable for their own content knowledge, the requirement is limited to teachers in the lowest-performing schools—and thus will likely overlook many teachers with poor content knowledge in higher-performing schools. In addition, it remains unclear how schools and districts will use this information.

What are the strengths and limitations of Ohio’s accountability system?

Strengths

Transparent data. In addition to the school report card issued by all states, Ohio’s creation of the web-based interactive Local Report Card (iLRC) steps up the level of transparency typically found in most state reporting systems. The system easily allows comparison among schools and districts, a feature not found in every state.

Comprehensive picture of school performance. With the use of the State Indicators, the Performance Index, the Value-Added Measure, and AYP, Ohio provides a more comprehensive picture of student learning and growth than states that simply report on student proficiency. (Though Ohio has distorted its AYP proficiency measures by allowing students below proficiency, but making “adequate” growth, to be counted as proficient. Further, only the AYP measure includes student subgroups.)

Value-added measure. Ohio should be recognized for its efforts to build a metric that not only shows the value added at the school level, but with further refinement may help to answer the question, “How much value does an individual teacher add to a student’s learning over the course of an academic year?”

Differentiated support. Rather than a “one size fits all” response to low-performing districts and schools, the Ohio *Differentiated Accountability* model attempts to differentiate support and consequences for low performers based on the *number* of their subgroups failing to make

adequate progress—rather than overall AYP status. That said, some view the model as an escape hatch from NCLB sanctions, not a model driving authentic change.

Limitations

Limited testing of essential content areas. Ohio only tests writing and social studies at the high school level. This can lead to curriculum narrowing, and students may be caught off guard when faced with these assessments in the later grades. (It also limits the number of teachers whose effectiveness can be analyzed based on student test scores.) At minimum, the dearth of writing assessments may be addressed as new assessments aligned with the Common Core standards come online in 2014-15.

Complex performance-rating mechanism. While Ohio provides a comprehensive look at school performance, the various performance indicators can be confusing at the local level. The system employs different proficiency passing rates for the State Indicators and for AYP, meaning that the same school can pass one but not the other, and/or that the overall performance rating can reflect one but not the other. As a district superintendent described, “The report card ratings are very confusing and convoluted because of multiple measures,” despite the state’s transparency in reporting all indicators.

Weak outline of sanctions. Ohio traditionally adheres to the NCLB outline of sanctions for low performers—meaning that its own school and district performance designations carry little real weight—but even then it allows schools and districts much leeway in meeting the federal requirements. The state recently took steps to strengthen its accountability system through H.B. 153, but this legislation only addresses the lowest-performing 5 percent of schools, leaving all other struggling schools to fend for themselves. Certainly the capacity and will of any state department to turn around low-performing schools is questionable, but all states (not just Ohio) should think carefully about the role of the state department of education, district, business communities, non-profits and other entities in supporting a wider band of struggling schools. The state and district need not be the only players here.

Final Word

Ohio’s accountability system provides a comprehensive picture of school performance. Unfortunately, the state does little to put that information to good use. While it made great strides in strengthening school and individual

accountability with its passage of H.B. 153, the state still has far to go to ensure that the new measures are enacted with fidelity.

Information on Ohio's education-accountability system was primarily drawn from interviews with state representatives, district representatives, and local stakeholders, as well as from the Ohio Department of Education website at www.ode.state.oh.us. Additional information was drawn from the National Council on Teacher Quality's *2011 State Teacher Policy Yearbook*.

TEXAS

Overview

Texas boasts a rich history of accountability for its schools and students. The Lone Star State first introduced its school and district accountability system in 1993 with the passage of Senate Bill 7. Since that time, the state legislature has demanded more of students via new performance indicators, introduced more tests at more grade levels, and established additional accountability measures related to data and financial integrity.

Beyond school and district accountability policies, Texas also demands much from individual students. Grade-promotion policies, high school graduation requirements, “no pass, no play” provisions, and automatic college-admission entry for high achievers give students a stake in their own performance. (Unfortunately, the same cannot be said of teachers and administrators—the state does little to incentivize or support their performance on the job.)

Texas’s accountability system, however, suffers from two faults. First, in terms of district and school accountability, the system has perhaps grown *too* complex; districts report some difficulty in communicating clearly to stakeholders and focusing on the data elements most useful for school-improvement efforts.⁷⁹ Second, and more importantly, stakeholders have noted that the state drive for accountability at all levels—district, school, and individuals—has lost some steam over the last two decades, resulting in a weakening of some provisions and a lack of follow-through on others.

Texas’s school and district accountability system, however, is due for an extreme makeover. House Bill 3, passed during the 2009 legislative session and tweaked in 2011, charged the Texas Education Agency (TEA) with developing a new accountability system that overhauls the state’s standards and assessments, incorporates new measures of college and career readiness among the system’s performance indicators, and revamps the state’s accountability requirements and structure. (H.B. 3 does not alter provisions for individual students and adults in the system.)

⁷⁹ The system’s complexity is due in part to its structure. In addition to the accountability provisions for its district and charter schools, Texas operates a separate accountability system for its alternative-education campuses. In addition, a Performance-Based Monitoring Analysis System (PBMAS) evaluates specific programs in districts and charter schools in areas such as bilingual education/ESL, career and technology education, special education, and certain Title programs under the No Child Left Behind Act (NCLB). The state also manages a financial accountability system, the Financial Integrity Rating System of Texas (FIRST), which is designed to monitor school districts’ management of financial resources. All of these systems operate independently of the federal accountability system.

The TEA has already begun the process of developing a new state accountability structure for implementation in 2012-13. The system will be based on the new standards and assessments introduced in the beginning of the 2011-12 school year. While H.B. 3 lays the potential foundation for one of the most progressive accountability systems in the nation, the final structure of the system will be determined incrementally throughout the implementation process—and thus could either fulfill or fall short of that potential. During the transition from the prior assessment system to the new system, the state will suspend district and school accountability ratings while new student performance standards are set and the new accountability system is developed.

Below, we map Texas's progress against six key components of strong state accountability systems.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

Texas has fielded statewide assessments to measure student learning since its introduction of the Texas Assessment of Basic Skills (TABS) test in 1979 (though assessments were not tied to an accountability system until 1993). The state overhauled its assessment system a number of times before establishing the Texas Assessment of Knowledge and Skills Test (TAKS) in 1999. This year (2011-12), it transitions to its fifth assessment system, the State of Texas Assessments of Academic Readiness (STAAR).

STAAR represents the state's transition to an assessment system based on more rigorous standards. In past years, the Texas Essential Knowledge and Skills (TEKS) (i.e., the state's established K-12 curriculum standards) have varied in quality across subjects; while the state could boast strong reading and science standards, its mathematics and U.S. History standards fell short.⁸⁰ As a result of H.B. 3, Texas introduced College and Career Readiness Standards (CCRS) for English language arts, mathematics, social sciences, and science incorporated them into the TEKS. (Texas is one of five states that have not adopted the Common Core State Standards.) All students in grades three through eight are tested in reading and math, with students in specific grades also tested in science, social studies, and writing. In high school, students take twelve end-of-course exams—three each in English, math, science, and social studies—with the tests counting as 15 percent of final course grades. Students must earn passing averages in each subject area to receive a diploma.

⁸⁰ In recent reviews conducted by the Thomas B. Fordham Institute, Texas received an A-minus for its reading standards, a B for its science standards, a C for its mathematics standards, and a D for its U.S. History standards. The state's U.S. History standards in particular have drawn much criticism for their politicized distortion of history. See *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, and *The State of State Science Standards 2012*, at www.standards.educationgadfly.net/.

The reading and math STAAR tests in grades three through eight are linked to performance expectations for high school English III and Algebra II end-of-course assessments. As a result, the new tests themselves will indicate academic growth as well as proficiency, so that educators and parents can know if their students are on track to graduate prepared for college or career. Texas would do well to set more rigorous proficiency cut scores on its STAAR tests than it had for the TAKS tests: A 2011 analysis found that the state's cut scores for its TAKS assessments in fourth- and eighth-grade reading and eighth-grade math all equated to the *below basic* level on the National Assessment of Educational Progress (NAEP). Its fourth-grade math cut score equated to NAEP's *basic* level.⁸¹

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Texas provides a wealth of data on its state education agency website, disaggregated by school, district, and subgroup. The Academic Excellence Indicator System presents data on the percentages of students who Met Standard (i.e., achieved proficiency) in each grade, disaggregated by subgroup, for the current and previous year. (The state does not, however, report data for students who fail to reach proficiency, and it only reports data for those who achieved Commended Performance—i.e., surpassed proficiency—for all grades combined.) Also disaggregated by subgroup are test participation rates; English language learner progress indicators; attendance, dropout, and four- and five-year completion rates; AP/IB and SAT/ACT participation and passing rates; student and staff demographics; teacher salaries, years of experience, and program assignments; and annual operating expenditures. Unfortunately, the database does not allow users to make comparisons among those groups within the tool itself.

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

Texas is in the process of revamping its accountability system, but it currently rates its schools and districts based on student proficiency (in reading, writing, math, social studies, and science) and on indicators of completion rates and annual dropout rates. Subgroups included in the ratings are white, African American, Hispanic, and economically disadvantaged students, as well as English language learners. Texas does not currently include a growth measure in the system. A previous growth measure, known as the Texas Projection Measure (TPM), was discontinued in 2011 due to criticism that the measure artificially inflated school and district performance: TPM allowed students who failed the

⁸¹ National Center for Education Statistics, *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (Washington, D.C.: U.S. Department of Education, August 2011), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458>.

state assessment in a given year to be counted as having passed it if the calculation showed that they would be expected to pass in a future year. House Bill 3, however, requires that the new state accountability system include a true growth measure, not a projected growth measure.

Based on the indicators noted above, schools and districts are placed into one of four categories:

- Exemplary
- Recognized
- Academically Acceptable
- Academically Unacceptable

As noted above, Texas reports the number of students who Met Standard as well as the number of students who achieved Commended Performance. For a school to receive the Recognized designation, 15 percent of its student body, as well as 15 percent of its low-income students, must score at the Commended level in reading and math. To receive the Exemplary designation, schools must report that 25 percent of all students and of all low-income students are Commended in those two subject areas. It is expected that the designation and naming of categories will be significantly different under the new accountability system.

Unfortunately, the state does provide for a number of exceptions by which schools and districts can improve their ratings. For example, under certain circumstances, a school can raise its rating one category by demonstrating “required improvement,” a degree of growth outlined by the state. According to one stakeholder, such exceptions have been “the most serious problem with the system over the last three years” because they have “taken the heat off at the local level” and contributed to a flattening of improvement.

Separate from the state accountability system is Texas’s system for measuring and reporting AYP progress under NCLB. The state and federal systems in Texas employ different passing standards, proficiency targets, sanctions, and interventions. Under Texas’s yet-to-be-crafted accountability system, it is likely that the federal system will continue to operate parallel to the state system unless the reauthorization of ESEA renders their integration more practical.

4) A system of rewards and consequences to drive improvement at the school and district levels

Rewards

Texas offers no concrete incentives to promote school improvement, relying instead on performance-based school and district labeling to drive improvement

efforts. Beyond its basic performance categories, the TEA operates the Gold Performance Acknowledgement Criteria Program to recognize schools for performance on categories not included in the state rating system. The categories in this program include non-traditional indicators such as the following:

- Advanced course/dual enrollment completion
- Advanced Placement/International Baccalaureate results
- Attendance rate
- Commended performance (each core content area)
- Comparable improvement (reading ELA/mathematics)
- SAT/ACT results
- Meeting college readiness standard (high school ELA/mathematics)

This designation, however, is simply a public acknowledgment and offers no tangible reward.

Sanctions

Though a proud “local control” state, Texas outlines an intensive state intervention strategy for persistently low-performing schools and districts, including the option to close either when necessary. According to one stakeholder, however, Texas only takes advantage of its most serious intervention options when a school or district is “outrageously bad.” This lack of initiative, coupled with the inflation in performance ratings allowed by the accountability rating system, results in minimal pressure for schools or districts to improve.

DISTRICT SANCTIONS

District interventions are determined by accreditation status. Accreditation assignments (which are separate from academic accountability ratings) are based on a number of indicators primarily related to data quality/integrity, the district’s academic performance, and the district’s rating under the state’s financial accountability system. The four accreditation labels are as follows:

- Accredited
- Accredited-Warned
- Accredited-Probation
- Not Accredited-Revoked

The commissioner has a number of options at his/her discretion to address districts in all but the top level of accreditation status, including the appointment of a monitor, conservator, management team, or board of managers to ensure and oversee district-level support to campuses. The commissioner can also opt to close

a district based on a number of considerations, including multiple years of low performance, data quality issues, financial integrity issues, poor performance of any special programs, and the district's failure to comply with required improvement activities. (In the case of district closure, the most recent occurred in July 2011, whereby a neighboring district annexed the closed district.)

SCHOOL SANCTIONS

Texas's accountability system emphasizes the district's role in providing support to low-performing campuses. Once a school is designated Academically Unacceptable, the district must form a Campus Intervention Team that includes at least one member outside the district to help assess the school's needs and guide the construction of a School Improvement Plan. If a school does not improve its accountability designation the next year, it enters a "multi-year low-performing" category, with interventions and sanctions increasing in severity as each year of low performance continues.

Once a school has been rated unacceptable for two consecutive years, the commissioner must order campus reconstitution. Reconstitution entails the removal or reassignment of some or all of the school's instructional personnel (with some discretion allowed for actions that may have already occurred on the part of the district), as well as resubmission and approval of a redrafted School Improvement Plan.

If a school receives an unacceptable rating for a third consecutive year, the commissioner can apply one of three "ultimate sanctions": repurposing, alternative management, or closure.⁸²

Support

Texas has a bifurcated system of support for low-performing schools and districts due to its separate state and federal accountability structures. On one hand, the state system of support seems to be underfunded for state-identified low performers. The state provides external technical assistance providers to serve on Campus Intervention Teams, but beyond that, the state provides no additional support other than targeted assistance grant programs tied to particular initiatives.

⁸² A 2010 Fordham Institute study found that, in Texas, 3 percent of all low-performing district schools and 11 percent of all low-performing charter schools closed between 2003-04 and 2008-09. The study did not report whether these closures were due to low performance or for other reasons. See David Stuit, *Are Bad Schools Immortal? Turnaround and Shutdowns in Both Charter and District Sectors* (Washington, D.C.: Thomas B. Fordham Institute, 2010), <http://www.edexcellence.net/publications-issues/publications/are-bad-schools-immortal.html>.

Texas does have an extensive system of regional education service centers (ESCs), and by law the ESCs are required to provide professional development and technical assistance to low-performing schools in their regions.⁸³ The ESCs vary, however, in terms of their size and available resources. Currently, Texas operates a Turnaround Center at one of the ESCs to provide technical assistance to districts with Academically Unacceptable schools. But funding for this program was cut in the latest legislative session, and it will cease to operate in spring 2012.

On the other hand, because Texas is a high-poverty state and receives a large federal Title I allocation, the support system for schools in improvement under NCLB is well funded. The TEA has established a partnership with one of its twenty ESCs to house the technical assistance function for schools in improvement under NCLB. This School Improvement Resource Center relies largely on a cadre of technical-assistance providers to work directly with a school's principal to conduct a comprehensive needs assessment, develop a school improvement plan, and implement the activities in the plan. Additionally, the School Improvement Resource Center provides professional development, data-analysis tools and resources, and support to the schools in improvement under NCLB. (Recently, the School Improvement Resource Center began expanding its technical assistance to work at the district level as well as the school level.)

In response to complaints from districts, the state has made an effort to streamline and combine requirements and support for schools in improvement under both the federal and state accountability systems. For example, a school identified for improvement under both the federal and state accountability systems can submit a single consolidated improvement plan rather than two separate plans. Additionally, external support providers can address both federal and state improvement requirements, thus reducing the number of required external providers working with the school. Whether the new accountability system can better integrate and balance the systems of support is yet to be seen.

5) A system of rewards and consequences to drive improvement at the individual student level

While Texas has in place a number of rewards and consequences to drive individual students to perform at high levels, these policies are not always adequately enforced or supported. Since 1999, the state has required through its Student Success Initiative (SSI) program that fifth and eighth graders pass end-of-

⁸³ Texas has a system of twenty ESCs that work under the direction of and in coordination with the state education agency. ESCs provide services to assist districts and schools in improving student achievement and operating more economically and efficiently. While a board of directors governs each ESC, the commissioner of education participates in the selection of and approves the hiring of executive directors; annually evaluates each executive director's performance; and approves each center's operating budget. For more information, see the Texas System of Education Service Centers website at <http://www.texasresc.net/default.htm>.

year reading and math assessments in order to be promoted to the next grade. While a strong requirement on paper, this policy has weakened since its original enactment: At first, third graders were also required to pass end-of-year tests, but that requirement was lifted in 2009. In addition, students can circumvent the requirement if a “grade placement committee”—comprised of parents, teachers, and administrators—determines that the student is likely to perform at grade level after additional instruction. The outcome, as one stakeholder put it, is that “too many schools abuse the process, and as a result it’s been watered down and no longer holds the same ‘pressure’ that it did at [its] inception.” Finally, the funding for schools with large numbers of students not meeting the SSI standards has been drastically reduced recently, slashed by 92 percent in June 2011.⁸⁴

In addition to end-of-year tests, high school students take twelve end-of-course exams—three each in English, math, science, and social studies—with the tests counting as 15 percent of final course grades. Students must earn passing averages in each subject area to receive a diploma.

Texas was the first state to adopt a “no pass, no play” rule, which it did in 1984. The law prohibits students who fail any classes from participating in sports and other extracurricular activities for a three-week period. The original law required a suspension period of six weeks, but that was reduced to three weeks in 1995. (This sanction is more than other states have for students, but clearly three weeks provides little time for a student to improve his lackluster performance.)

Finally, Texas provides automatic college admission into a state college or university for those high school students who graduate in the top ten percent of their class.

Beyond consequences and rewards, Texas does support students in failing schools, at least in theory, by providing a number of outlets for them. Nearly every district in the state allows both inter- and intra-district transfers. In addition, the state established the Public Education Grant (PEG) program in 1995; it allows any student in a low-performing school to transfer to a higher-performing school in the same district or another district.⁸⁵

But again, while the state allows for such transfers, it does little to encourage students to take advantage of them. “It’s really choice in name only,” according to one stakeholder. In 2007-08 for example, only 220,000 students attended schools

⁸⁴ Thao Nguyen, “Texas merit pay plan for teachers among programs slashed by legislatures,” *Dallas Morning News*, June 24, 2011, <http://www.dallasnews.com/news/politics/texas-legislature/headlines/20110623-texas-merit-pay-plan-for-teachers-among-programs-slashed-by-legislature-.ece>.

⁸⁵ Low-performing PEG schools are defined as those in which 50 percent of students failed any subject included in the state assessment system in any two of the last three years, and those rated Academically Unacceptable in any one of the last three years.

other than their home schools—either through transfers, PEG, charter schools, magnet schools, or NCLB provisions—amounting to only 5 percent of all Texas students.⁸⁶

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

Texas has in place some policies to drive teacher and administrator performance, but often these provisions are either not adequately upheld, financially supported, or statewide in nature.

Texas's instrument for evaluating teachers is the Professional Development and Appraisal System (PDAS). (While about 86 percent of all Texas districts use PDAS to evaluate their teachers, districts may opt out of the system and install their own evaluation metrics.) PDAS does not require that teacher evaluations incorporate objective measures of student learning—only “observable” student performance. Texas does have in place a provision for the state to develop an evaluation model that includes student achievement data. This policy was passed three legislative sessions ago, but the state has not yet acted on it, and a model has not yet been developed. Even then, districts would have the option to develop their own models or opt for the state-developed model.

For now, PDAS encompasses eight domains. Each is scored independently for each teacher, using the following ratings: exceeds expectations, proficient, below expectations, and unsatisfactory. A teacher scoring “unsatisfactory” on any one of the domains, or “below expectations” on any two, is placed on an intervention plan developed by the appraiser. In theory, if a teacher does not meet the conditions for improvement spelled out in his or her intervention plan, he or she may be eligible for dismissal—though Texas's language surrounding this provision is somewhat murky, and teachers can appeal dismissal multiple times.

Texas requires that only some teachers be evaluated annually. While new teachers must be evaluated each year, non-probationary teachers (i.e., those with tenure) can circumvent this requirement by receiving “proficient” ratings on their most recent evaluations and agreeing in writing to be evaluated at least once every five years. Tenure is conferred virtually automatically after three years. In times of layoffs, however, Texas requires that “teacher appraisals” be the chief criterion for deciding which teachers are removed; in other words, seniority is not allowed to drive the decision.⁸⁷

⁸⁶ “Choice in Texas Public Schools,” Coalition for Public Schools, accessed January 9, 2012, http://www.coalition4publicschools.org/coalition/documents/choice_fact_sheet.pdf

⁸⁷ An analysis by the Texas Education Agency found that 73 percent of districts used evaluation results to inform removal decisions in 2010-11, while just 7 percent used them to inform compensation. See “Uses of

Texas requires that districts adhere to a minimum salary schedule based on seniority. Still, the state provides some support for merit-based teacher pay. The District Awards for Teacher Excellence (DATE) provides non-competitive grants to districts to distribute among its highest-performing educators and personnel. At least 60 percent of the funds must be distributed based on objective evidence of increased student performance. State funding for DATE, however, was slashed by almost 90 percent in Texas's 2011 legislative session, limiting its reach and impact.

Texas does support differential pay for individuals teaching certain subjects and in high-needs schools. These teachers are eligible for an additional \$5,000 annual stipend.

For administrators, Texas requires annual evaluations but does not require that they incorporate objective measures of student performance, nor does it tie them to employment decisions.⁸⁸

What are the strengths and limitations of Texas' accountability system?

Strengths

Robust data collection. Texas does a good job of providing districts and schools with data on multiple indicators of student learning. In addition to student proficiency, Texas reports completion rates and annual dropout rates, and disaggregates nearly all data. Texas has long been a go-to state for education research due to its rich data sets over multiple years.

Outline of intensive sanctions. Texas delineates a number of sanctions for low-performing schools, laying the groundwork for a strong accountability system—much more than can be said of most states. Still, the state has not always worked to overcome the political “push back” that occurs when it intervenes at the local level, despite a strong business coalition that urges the state to take responsibility for its low-performing schools. Thus, in recent years the state has declined to take advantage of all its options except when it comes to the very worst schools.

A focus on career and college readiness. While it hasn't yet taken full effect, H.B. 3 has the potential to usher in a new era of district, school, and student/teacher accountability in Texas. The recently adopted College and

teacher evaluation results,” Texas Education Agency, August 29, 2011, www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147502760&libID=2147502754.

⁸⁸ According to the TEA, 66 percent of districts used principal evaluations to inform removal decisions in 2010-11, while just 14 percent used them to inform compensation. See “Uses of principal evaluation results,” Texas Education Agency, August 29, 2011, www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147502751&libID=2147502745.

Career Readiness Standards (CCRS) align PK-16 and higher education curricula in English language arts, math, social sciences, and science. The state boasts an ambitious PK-16 agenda including a legislatively mandated partnership between the TEA and the Texas Higher Education Coordinating Board to increase college readiness and college-going rates while reducing higher-education attrition. The state is in the process of establishing and implementing an accountability system for its state institutions of higher education. A risky venture, to be sure, but one to watch.

A focus on student accountability. Texas has in place numerous provisions that place students in the driver's seat of their own education, from grade promotion policies and high school graduation requirements to "no pass, no play" regulations and automatic college admission for high performers. But like many areas in Texas's accountability system, these policies are not always adequately supported or enforced at the state level, and some have been weakened over time.

Limitations

Complex system of accountability. While Texas should be commended for its comprehensive accountability system, the amount of data it collects can be unwieldy. The system does not distinguish among less and more important data, and does not clearly communicate to stakeholders which information matters most.

Dual federal and state accountability systems. Districts report the two separate accountability systems to be in conflict, as they employ different passing standards, proficiency targets, sanctions, and interventions. While the state has attempted to reduce conflicting and duplicative requirements for schools in improvement under both, the two systems remain burdensome.

Few resources for low-performing schools or incentives for high-performing schools. While some resources and support are provided through the federal accountability system, districts perceive the state accountability system and its associated designations and sanctions as the system "with the teeth." The latter clearly outlines sanctions for low-performing schools, but lacks a clear system of support for those schools in improvement. In addition, it provides few concrete incentives for schools to achieve at high levels.

Few teacher consequences or incentives. Texas does not mandate that objective measures of student achievement be included in teacher or administrator evaluations, nor does it clearly tie evaluations to employment

decisions, except in times of teacher layoffs. In addition, with the drastic slashing of its teacher merit-pay grant program, highly effective teachers receive few concrete incentives.

Final Word

While Texas's accountability movement may have lost some steam in recent years, the state deserves credit for setting a precedent of accountability for the entire nation. (And the nation will again be watching as the Lone Star State joins a small band of brothers in going it alone with its own standards and tests in lieu of the Common Core.) In addition, the state's passage of H.B. 3 signals a renewed commitment to a robust and stalwart system of accountability for schools and district—at least in theory, as final details of the accountability structure have not yet been determined. In the meantime, Texas would do well to bolster its accountability provisions for the adults in the system, and to ensure that student accountability policies are implemented with greater fidelity.

Information on Texas's education-accountability system was primarily drawn from interviews with state representatives, district representatives, and local stakeholders, as well as from the Texas Education Agency website at www.tea.state.tx.us. Additional information was drawn from the ESC Region XIII website at <http://www5.esc13.net/pdas/> and from the National Council on Teacher Quality's *2011 State Teacher Policy Yearbook*.